THE UNIVERSITY *of* EDINBURGH

THE UNIVERSITY
of EDINBURGH

A structural and synthetic biology study of bacterial microcompartments

Laura Tuck

Submitted for the degree of Doctor of Philosophy

University of Edinburgh

April 2018

Publications

Tuck, L.R., Altenbach, K., Ang, T.F., Crawshaw, A.D., Campopiano, D.J., Clarke, D.J., and Marles-Wright, J. (2016). Insight into Coenzyme A cofactor binding and the mechanism of acyl-transfer in an acylating aldehyde dehydrogenase from Clostridium phytofermentans. Sci. Rep. *6*, 22108.

Liu, W., **Tuck, L.R**., Wright, J.M., and Cai, Y. (2017). Using purified tyrosine site-specific recombinases in vitro to rapidly construct and diversify metabolic pathways. In Methods in Molecular Biology, pp. 285–302.

Liu, W., Luo, Z., Wang, Y., Pham, N.T., **Tuck, L**., Pérez-Pi, I., Liu, L., Shen, Y., French, C., Auer, M., et al. (2018). Rapid pathway prototyping and engineering using in vitro and in vivo synthetic genome SCRaMbLE-in methods. Nat. Commun. *9*, 1936.

Conference presentations and poster sessions

Dec 2016 – Oral presentation at ICB/IQB Postgraduate Symposium

Aug 2016 – Invited speaker at 'Structural aspects of synthetic biology systems' symposium, University of Minnesota, USA

Feb 2016 – 'Applied synthetic biology in Europe' conference, Lisbon, Portugal. **Poster**

Aug 2015 – CCP4 Northern Protein Structure Workshop, Carlisle, UK

Jun 2015 – SULSA 2015 Synthetic Biology Meeting, Dundee, UK. **Organiser**

Jun 2014 – European Congress on Biotechnolgy, Edinburgh, UK. **Volunteer**

Jun 2014 – SULSA 2014 Synthetic Biology Meeting, Edinburgh, UK. **Poster**

Professional Internships for PhD Students (PIPS) placement

As the Events and Conference Co-ordinator for the Scottish Universities Life Sciences Alliance (SULSA), my main goal was to organize the annual SULSA Young Scientists Event (YSE), which is aimed at all SULSA PhD students and is based around a general theme on which guest speakers are invited to give talks. I was responsible for all aspects of the event, from identifying a suitable theme, deciding a venue and inviting speakers and students. My major achievement was booking Ben Goldacre as a guest speaker, and although he was not able to attend the event on the day, his talk was re-scheduled as was opened out to members of the public as well as YSE attendees with around 300 people in attendance. The excellent talk brought great exposure for the YSE and the theme of 'Science Portrayal in the Media'. The event itself was also successful, with brilliant talks and interactive sessions from Alison Gray, Joanna Young, Sara Shinton and Tom Pringle; a conference dinner and a team building event, which was enjoyed by all the attendees. The placement was a fantastic opportunity for network, social outreach and for my personal development.

Table of Contents

Table of Figures *8*

Table of Tables *10*

Acknowledgements *11*

Abbreviation list *13*

Declaration *15*

Lay Summary *16*

Abstract *19*

1 Introduction *21*

1.1 Bacterial microcompartments **21**

1.1.1 α and β–Carboxysomes 23

1.1.2 Metabolosomes 23

1.2 BMC metabolic pathways **26**

1.2.1 α-carboxysome and β-carboxysome metabolism 26

1.2.2 1,2 – propanediol metabolism 27

1.2.3 Ethanolamine metabolism 29

1.2.4 Ethanol metabolism 30

1.3 Glycyl-radical enzymes **30**

1.4 Overview of BMC protein shell structure **31**

1.4.1 Hexameric domains (BMC-H) 33

1.4.2 Tandem hexameric domains (BMC-T) 34

1.4.3 Pentameric domains (BMC-P) 34

1.4.4 Microcompartment shell assembly 34

1.5 Bacterial microcompartment enzyme-shell targeting **35**

1.6 BMCs as tools for synthetic biology **36**

1.6.1 Bacterial microcompartments in bioengineering 36

1.7 Golden Gate cloning and Type II restriction enzymes **37**

1.7.1 Type II restriction endonucleases 38

1.7.2 CIDAR MoClo: Modular assembly standard 38

1.8 Aims and objectives **41**

2 Expression and purification of acylating aldehyde dehydrogenase enzymes from Clostridium phytofermentans bacterial microcompartment operon *43*

2.1 Introduction **43**

2.1.1 Overexpression of recombinant Cphy1178, Cphy1428 and Cphy2642 43

2.1.2 Purification of recombinant aldehyde dehydrogenases by anion exchange and size-exclusion chromatography 45

2.1.3 Summary and Discussion 53

2.2 Enzyme activity of acylating aldehyde dehydrogenase enzymes from Clostridium phytofermentans bacterial microcompartment operons **53**

2.2.1 Introduction 53

2.2.2 Preliminary NAD+ turnover assay set-up 54

2.2.3 Enzyme assay set-up 54
2.2.4 Enzyme assay results 55
2.2.5 Expression and purification of the active site aldehyde dehydrogenase variant Cphy1178(20-462) (C269A and H387A) 59
2.2.6 Summary and discussion 61

3 Crystal structure of an acylating aldehyde dehydrogenase from Clostridium phytofermentans, with Co-enzyme A and NAD+ in the active site 62

3.1. Introduction 62
3.1.1 Crystallisation of Cphy1178(20-462) and active site variants 66
3.1.2 Structure of Cphy1178$_{(20-462)}$ 70
3.1.3 Structure of Cphy1178(20-462) C269A and H387A variants 71
3.1.4 Co-factor binding 72
3.1.5 Proposed mechanism of action of Cphy1178(20-462) 77
3.1.6 Structure of Cphy1178(20-462) E357A variant 80
3.1.7 Summary and discussion 82

4 Bacterial microcompartments as tools for synthetic biology 84

4.1 Introduction 84

4.2 Shell protein library construction using CIDAR MoClo toolkit 86
4.2.1 Introduction 86
4.2.2 Shell proteins in Clostridium phytofermentans fucose/rhamnose and Salmonella Pdu BMCs 87
4.2.3 Establishing the MoClo DNA assembly pipeline for the production of synthetic BMCs 90
4.2.4 Rational design of shell protein constructs 91
4.2.5 Summary and discussion 94

4.3 RBS + Cphy1178/STM2051 localisation sequence constructs 94
4.3.1 Introduction 94
4.3.2 RBS + Cphy1178/STM2051 localisation sequence construct design 95
4.3.3 RBS + Cphy1178/STM2051 localisation sequence plate reader assays 98
4.3.4 RBS + Cphy1178/STM2051 localisation sequence plate FACS analysis 103
4.3.5 Summary and discussion 103

4.4 EutM shell proteins 105
4.4.1 Introduction 105
4.4.2 pET28a-GG-RFP plasmid for one-pot cloning of ORFs for overexpression 105
4.4.3 Construct expression and purification by ammonium sulfhate precipitation 106
4.4.4 Thin-section electron microscopy of EutM shell proteins 111
4.4.5 Summary and discussion 116

5 Discussion 117

5.1 Expression and purification of acylating aldehyde dehydrogenase enzymes . 117
5.1.2 Activity of aldehyde dehydrogenase Cphy1178 against short chain fatty-aldehydes 117
5.1.3 Crystal structures of aldehyde dehydrogenase Cphy1178 with NAD^+ and Co-enzyme A in the nucleotide binding domain 118

5.2 Building synthetic microcompartments 119
5.2.1 Bacterial microcompartment shell protein constructs 119
5.2.2 Forward engineering of RBS + aldehyde dehydrogenase localisation sequences to study downstream protein expression 120
5.2.3 EutM shell proteins – expression and thin-section electron microscopy 121

5.3 Future work 122
5.3.1 Propionaldehyde dehydrogenase structure and biochemistry 122

5.3.2 Bacterial microcompartments as tools for synthetic biology 123

6 Materials and Methods 124

6.1 Aldehyde dehydrogenase enzyme cloning 124
6.1.1 Full-length and truncated protein cloning 124
6.1.2 Site-directed mutagenesis of Cphy1178$_{(20-462)}$ active site residues 127

6.2 Aldehyde dehydrogenase enzyme expression, purification and activity assays 128
6.2.1 Aldehyde dehydrogenase expression 128
6.2.2 Aldehyde dehydrogenase purification 128
6.2.3 Aldehyde dehydrogenase activity assays 129

6.3 Crystallisation/Optimisation 131
6.3.1 Protein crystallisation screening and optimisation 131
6.3.2 Crystal soaking and mounting 131
6.3.3 X-ray diffraction data collection and processing 132
6.3.4 Crystallographic model building and refinement 132

6.4 Cidar MoClo cloning 134
6.4.1 Insertion of shell protein constructs into Level 0 vector 134
6.4.2 Insertion of terminator tags into MoClo toolkit 137

6.5 RBS + localisation sequence design and cloning 138
6.5.1 RBS + localisation sequence design 138
6.5.2 RBS+localisation sequence cloning 138
6.5.3 RBS + localisation sequence plate reader assays 142
6.5.4 RBS+ localisation sequence flow cytometry 142

6.6 EutM cloning, protein expression and purification and thin- section electron microscopy 142
6.6.1 Cloning EutM into Golden-Gate modified pET28a 142
6.6.2 Ammonium sulfate precipitation of *D. psychrophila* and *P. hadalis* EutM proteins .. 144
6.6.3 Thin-section electron microscopy of EutM constructs 144

7 Appendix 145

7.1 Calibration of HiLoad 16/600 Superdex 200pg Size Exclusion Chromatography column 145

8 Bibliography 147

Table of Figures

Figure 1.1. Transmission electron micrographs of BMCs 22
Figure 1.2. Schematic representation of simplified metabolic pathways within BMCs. 27
Figure 1.3. Genes associated with the fucose/rhamnose breakdown pathway in.....29
Clostridium phytofermentans 29
Figure 1.4. A cartoon representation of the crystal structure of the EutM shell protein from Clostridium difficile 32
Figure 1.5. A representation of the three main types of bacterial microcompartment shell proteins and their pfam domains 33
Figure 1.6 SBOL generated figures to show the assembly of the MoClo transcriptional units. 40
Figure 2.2 Expression tests of BMC-associated aldehyde dehydrogenases from.....44
C. phytofermentans produced recombinantly in *E. coli* BL21 (DE3) cells. 44
Figure 2.3 Anion exchange purification of recombinant Cphy1178. 47
Figure 2.4 Size-exclusion chromatography purification of recombinant Cphy1178. 48
Figure 2.5 Anion exchange purification of recombinant Cphy1428. 49
Figure 2.6 Size-exclusion chromatography purification of recombinant Cphy1428. 50
Figure 2.7 Anion exchange purification of recombinant Cphy2642. 51
Figure 2.8 Size-exclusion chromatography purification of recombinant Cphy2642. 52
Figure 2.9 Activity of Cphy1178 against various aldehyde substrates. 57
Figure 2.9 Activity of Cphy1178 against various aldehyde substrates. 58
Figure 2.10 Size-exclusion chromatography purification of recombinant Cphy1178 (C269A) variant. 60
Figure 3.1 Multiple sequence alignment of aldehyde dehydrogenase enzymes used in this study and close homologues. 66
Figure 3.2 Images of bi-pyramidal Cphy1178$_{(20-462)}$ and variant crystals. .....**Error! Bookmark not defined.**
Figure 3.3 Structure of Cphy1178$_{(20-462)}$ 71
Figure 3.4 Surface views of cofactor binding to Cphy1178$_{(20-462)}$. 73
Figure 3.5 Stick representation of NAD+ binding in the cofactor binding site. 74
Figure 3.6. Wall-eyed stereo view of electron density maps of bound NAD+ and CoA cofactors. 75
Figure 3.7 Wall-eyed stereo view of cofactor binding within the nucleotide- binding domain of Cphy1178$_{20-462}$. 77
Figure 3.8 Proposed catalytic mechanism of Cphy1178. 79
Figure 3.9 Architecture of cofactor binding to Cphy1178(20-462). 81
Figure 3.10. Electron density map of potential bound CoA cofactor 82
Figure 3.11 Scheme for Clostridium phytofermentans fucose/rhamnose utilisation pathway. 83
Figure 4.1. The synthetic biology workflow (Myers et al., 2017).....85
Figure 4.2 Multiple sequence alignment of the BMC-H type shell proteins used in this study, aligned against the EutM homologue from *C. difficile.* 88
Figure 4.3 Multiple sequence alignment of the BMC-P type shell proteins used in this study, aligned against the EutN homologue from E. coli. 89

Figure 4.4 Multiple sequence alignment of the BMC-T-type shell proteins used in this study, aligned against the PduT homologue from *Citrobacter freundii.* 90
Figure 4.5 SBOL images of shell protein constructs, designed with Benchling software and generated using the modular cloning (MoClo) toolkit, in a one-pot assembly method. 92
Figure 4.6 SBOL representation of RBS+loc construct transcriptional unit with fluorescent reporter gene.....95
Figure 4.7 The design, build, test, learn, use synthetic biology workflow cycle as related to the RBS+loc sequence constructs. 97
Figure 4.8 Time-course OD600 growth curves using construct TU1 from Table 4.398
Figure 4.9 Time-course FL485-538 curves using construct TU1 from Table 4.3 99
Figure 4.10 Time-course FL485-538/OD600 growth curves using construct TU1 from Table 4.3 100
Figure 4.11 Time-course OD600, FL485-538 and FL485-538/OD600 growth curves using construct TU1 from Table 4.3 measured at IPTG concentration of 0.5 mM..101
Figure 4.12 Histograms of the RBS+loc constructs at 0 mM, 0.5 mM and 1 mM IPTG concentration.....102
Figure 4.13 Histograms of the geometric means from FACS sorting analysis for Cphy1178 and STM2051 RBS+loc constructs 104
Figure 4.14 Plasmid map of the pET28a vector, engineered into a MoClo compatible protein expression vector 106
Figure 4.15 Anion exchange purification of recombinant EutM from *Desulfotalea*108
psychrophila. 108
Figure 4.16 Anion exchange purification of recombinant EutM from *Psychromonas hadalis*.....109
Figure 4.17 Ammonium sulfate precipitation of recombinant EutM from.....110
Desulfotalea psychrophila (Dp) and *Psychromonas hadali* (Ph).....110
Figure 4.18 Thin-section transmission electron micrograph images of overexpression of recombinant EutM protein from *Desulfotalea psychrophila* in *E. coli* cells 112
Figure 4.19 Thin-section transmission electron micrograph images of overexpression of recombinant EutM protein from Psychromonas hadalis in E. coli cells 113
Figure 4.20 Thin-section transmission electron micrograph images of overexpression of recombinant EutM protein from *Salmonella enterica* in *E. coli* cells.....114
Figure 4.21 Thin-section transmission electron micrograph images of overexpression of recombinant EutM protein from *Thauera linaloolentis* in E. coli cells.....115
Figure 1 Chromatographic separation and calibration curve for HiLoad 16/600 Superdex 200 pg column.....146

Table of Tables

Table 1.1 5' and 3' toolkit defined overhangs of the MoClo system 40
Table 2.1 Calculated parameters of the three aldehyde dehydrogenase enzymes.....45
Table 3.1 Amino acid sequence identity (%) between aldehyde dehydrogenases from different bacterial species.....63
Table 4.1 BMC shell proteins from *Clostridium phytofermentans* and.....87
Salmonella enterica used in this study. 87
Table 4.2 Rational design of shell protein level 1 transcriptional units. 93
Table 4.3 Transcriptional units generated with the RBS+loc constructs.....96
Table 6.1 Primers used for the aldehyde dehydrogenase cloning 126
Table 6.2 Primers used for shell protein to Level 0 cloning and.....135
Table 6.3 Level 0 Golden-Gate cloning reaction components and volumes.....136
Table 6.4 Level 0 Golden-Gate cloning digestion and ligation parameters.....136
Table 6.5 Forward and reverse primers for terminator tag cloning.....137
Table 6.6 Forward and reverse PCR primers for RBS + localisation sequence constructs.....139
Table 6.7 Level 1 Golden-Gate cloning reaction components and volumes 140
Table 6.8 Level 1 Golden-Gate cloning digestion and ligation parameters.....140
Table 6.9 Level 1 RBS+loc transcriptional unit constructs 141
Table 6.11 Primers for EutM constructs to be cloned into the amended Golden-Gate pET28a vector 143

Acknowledgements

I would like to give my sincere thanks to my supervisors, Dr Jon Marles-Wright and Dr Janice Bramham. Their support and mentorship has been invaluable to me throughout the past 4 years, particularly during the writing of this thesis.

Jon has been an excellent mentor, with his willingness to devote time and energy to helping me grow as a scientist, as well as teaching me invaluable lab-based skills that I will use throughout my future career. Despite Jon leaving for Newcastle University at the end of my third year, he remained committed to and engaged with my project, ensuring that everything stayed on track, and that I was supported by my replacement supervisor, Janice. I truly appreciate the support I was given by Janice and my second supervior, Professor Malcolm D. Walkinshaw, with particular thanks to Janice for her help with the revision of my thesis.

I would like to thank the BBSRC Eastbio Doctoral Training Programme for their funding and support, as well as the James Rennie Bequest for my conference travel funding.

My sincerest thanks go to all of my colleagues past and present, particularly Didi He, not only for her advice and support, but also her friendship (and all the delicious shared food). Many thanks to Dr. Kirsten Althenbach, Efrain Zarazua-Arvizu and Miguel Cueva.

I would like to thank all people in Swann level 3 labs, especially the scientists from EPPF including Dr. Liz Blackburn, Dr. Martin Wear and Dr. Matt Nowicki for their training and advice.

Particular thanks goes to all my amazing family, but especially my Mam, Dad, Claire and Sophie. I can't thank you enough for your love and support, throughout my whole

life, but particularly during the last few years. I am so lucky to have you and I truly appreciate everything you've done for me.

I've made some amazing friends in Edinburgh, so particular thanks to Dr. Joanna Strachan (seriously, thank you for everything Jo), Steph Hicks, Valdeko Kruusvee, Zac Boyd and Ben McCusker.

Last, but certainly not least I would like to thank Dr Elliott Chapman. I am more grateful to you than you can know.

Abbreviation list

1,2-PD	1,2-propanediol
A_{280}	Absorbance at 280 nm
A_{315}	Absorbance at 315 nm
AldDH	Aldehyde dehydrogenase
AFM	Atomic force microscopy
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
BMC	Bacterial micrcompartment
BME	2-mercaptoethanol
CA	Carbonic anhydrase
CCM	Carbon concentrating mechanism
CDS	Coding DNA sequence
CIDAR	Cross-disciplinary Integration of Design Automation Research
CO_2	Carbon dioxide
CoA	Coenzyme A
Cphy	*Clostridium phytofermentans*
DIC	Dissolved inorganic carbon
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
Dp	*Desulfotalea psychrophila*
E. coli	*Escherichia coli*
Etu	Ethanol utilisation microcompartment
Eut	Ethanolamine utilisation microcompartment
FACS	Fluorescence
FUC	Fucose/rhamnose activated cell sorting
GFP	Green fluorescent protein
GRE	Glycyl- radical enzyme
$HCO3^-$	Bicarbonate
iGEM	International genetically engineered machine
IPTG	Isopropyl β- D-thiogalactopyranosidase
kDa	Kilo Dalton
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LB medium	Lysogeny broth
MCS	Multiple cloning site
MoClo	Modular Cloning
MS	Mass spectrometry
MW	Molecular weight
MWCO	Molecular weight cut off
NAD +	Nicotinamide adenine dinucleotide

$NADP^+$	Nicotinamide adenine dinucleotide phosphate
NEB	New England Biosciences
ORF	Open reading frame
PCR	Polymerase chain reaction
PDB	Protein Data Bank
Pdu	Propanediol utilisation
pI	Isoelectric point
Ph	*Psychromonas hadalis*
RBS	Ribosome binding site
RFP	Red fluorescent protein
RMSDCα	Root-mean square deviation of Cα positions
RuBisCO	Ribulose 1,5 bisphosphate carboxylase/oxygenase
SBOL	Synthetic Biology Open Language
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
STM	*Salmonella enterica*
TCEP	Tris(2 carboxyethyl)phosphine hydrochloride
TEM	Transmission electron microscopy
Tl	*Thauera linaloolentis*
TU	Transcriptional unit
X-gal	5bromo-4-chloro-3-indolyl-β-D-galactopyranoside
YFP	Yellow fluorescent protein

Declaration

The work presented in this thesis was conducted solely by the author, and any contributions by other people have been clearly indicated and referenced. The work has not been submitted for any other degree or professional qualification.

Signed,

Laura Tuck

Lay Summary

For many years, the insides of bacterial cells were thought to be unstructured, with key cell components 'floating around' in the cytosol. However, around 60 years ago, small protein shells were discovered in a type of bacteria that get their energy from photosynthesis and were named carboxysomes. Carboxysomes were found to contain two enzymes that take carbon dioxide and 'fix' it into useful sugars: ribulose-1,5-bisphosphate carboxylase/oxygenase (known as RuBisCO) and carbonic anhydrase (CA). Alone, RuBisCO is a very poor enzyme, in that the conversion of CO2 to sugar is very slow. By localising the RuBisCO and carbonic anhydrase, which makes carbon dioxide from bicarbonate, together in the protein shell, the efficiency of RuBisCO is much increased.

Since the discovery of the carboxysomes, other types of protein shells containing enzymes have been found in a wide range of different bacteria, and these have been named bacterial microcompartments (BMCs). BMCs look similar to carboxysomes, but they work in the opposite direction: they break down sugars to provide the cells with energy. However, these reactions produce toxic chemicals, so the BMCs also protect the cell from these by keeping them inside their shell. One particular BMC is the fucose/rhamnose microcompartment in the soil bacterium Clostridium phytofermentans, which takes the sugars fucose and rhamnose and breaks them down into molecules used for the energy within the cell. One important enzyme in this pathway is called an aldehyde dehydrogenase; it uses two different molecules (NAD+ and Coenzyme A {CoA}) to turn an aldehyde (a component of fucose/rhamnose breakdown) into an acyl-CoA, which can be used to generate energy for the cell. Although the general structure of the aldehyde dehydrogenase had been solved by X-

ray crystallography, up until now it was unclear how this enzyme works. The first part of this project shows the crystal structure of the aldehyde dehydrogenase with the NAD+ and CoA cofactors in the active site; the latter for the first time. Along with data generated from an enzyme activity experiment, shows how this enzyme works, which is an exciting development in the field of BMC study.

The second part of the project is focused on synthetic biology. A common definition of synthetic biology is 'designing and constructing biological modules, biological systems and biological machines, or re-design of existing biological systems for useful purposes. Shell proteins of BMCs are described as 'modular', because they are made of three different protein types, two proteins that form hexagonal arrangements and a pentagon that can be put together in many different combinations to form the 'football-shaped' microcompartment. As BMCs are beneficial to the cell because they protect it from toxic molecules, it is possible to use the principals of synthetic biology to design and make synthetic BMCs, which can then be used to target different useful molecules inside them. This part of the project looks at the synthetic biology of BMCs from three different angles:

1. Can we use a specifically designed toolkit of biological parts to generate synthetic microcompartments using the protein shells from two different types of bacteria?
2. Can we take the protein sequence that directs the enzymes to the BMC, from the aldehyde dehydrogenase described in the earlier part of the project and use it to see if it affects the level of expression of a particular fluorescent protein?

3. Finally, can we take a particular type of shell protein called EutM, from four different bacteria, and overexpress it in E. coli to see what happens to the shell of the BMC

The answers to these three questions can help towards building useful synthetic bacterial microcompartments

Abstract

Bacterial microcompartments (BMCs) are proteinaceous metabolic compartments found in a wide range of bacteria, whose function it is to encapsulate pathways for the breakdown of various carbon sources, whilst retaining toxic and volatile intermediates formed from substrate breakdown. Examples of these metabolic processes are the 1,2-propanediol-breakdown pathway in Salmonella enterica (Pdu microcompartment), as well as the ethanolamine breakdown pathway in Clostridium difficile (Eut microcompartment). Some of the major challenges to exploiting BMCs as a tool in biotechnology are understanding how enzymes are targeted to microcompartments, as well as being able to engineer the protein shell of BMCs to make synthetic microcompartments that allow specific enzyme pathways to be targeted to their interior. Finally, the metabolic burden imposed by the production of large protein complexes requires a detailed knowledge of how the expression of these systems are controlled.

This project explores the structure and biochemistry of an essential BMC pathway enzyme, the acylating propionaldehyde dehydrogenase. With crystal structures of the enzyme with the cofactors in the cofactor binding site and biochemical data presented to confirm the enzyme's substrate. The project also focuses on the creation of synthetic biology tools to enable BMC engineering with a modular library of BMC shell protein parts; forward engineered ribosome binding sites (RBS) fused to BMC aldehyde dehydrogenase localisation sequences. The parts for this library were taken from the BMC loci found in Clostridium phytofermentans and Salmonella enterica. Using a synthetic biology toolkit will allow the rapid prototyping of BMC constructs for use in metabolic engineering. The shell protein parts were used to generate a number of

transcriptional units, to assess the effect of overexpression of individual BMC shell components on the morphology of BMCs and the effect these had on their host chassis. Different strength forward engineered RBS and localisation constructs have been designed to assess the possibility of controlling the levels of heterologous proteins targeted to the interior of microcompartment shell to allow metabolic engineering of encapsulated pathways. Along with looking at overexpression of a single shell protein, to assess viability of BMCs as scaffold-like structures, recombinant BMCs can be explored for their utility in bioengineering and their potential role in generating biofuels.

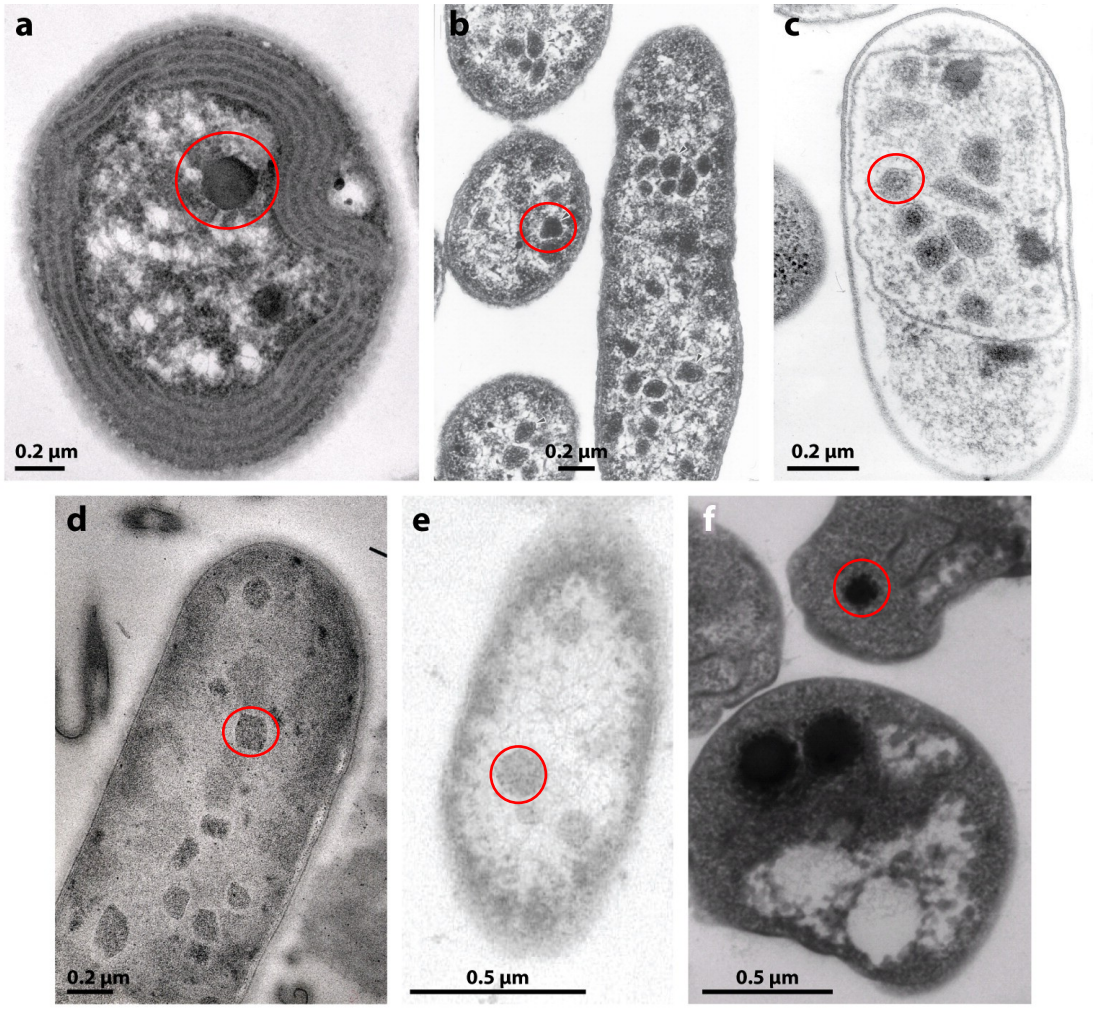
1 Introduction

1.1 Bacterial microcompartments

Bacteria have been used for many years as model organisms to study many different facets of biology. Most of the knowledge gleaned from basic molecular processes such as transcription and translation comes from initial studies into bacterial molecular biology. However, the bacterial cell was first thought to be unstructured and uncompartmentalised, unlike eukaryotic cells with their complex internal organisation (Gitai, 2005). This view was challenged around 50 years ago when bacterial microcompartments (BMCs) were first observed in the chemoautotrophic bacterium *Halothiobacillus neapolitanus* (Shively et al., 1973), by thin-section electron micrographs of the bacterium. They found that the compartments contained ribulose 1,5-bisphosphate, which is involved in carbon dioxide fixation, and proposed that the structures be named carboxysomes.

A BMC consists of a pseudo-icosahedral shaped protein shell, which is selectively permeable to substrates and products formed from either anabolic or catabolic reactions, and contains enzymes to metabolise these reactions, shown in Figure 1.1 (Kerfeld et al., 2010, 2018). This form of compartmentalisation has arisen as a way to enhance metabolic flux through multi-step pathways, and may also increase enzyme stability (Jakobson et al., 2017). The separation of the encapsulated enzymes and the bacterial cytosol is also thought to protect the cell from toxic volatile intermediates (Chen and Silver, 2012). There are two main types of BMCs, α- and β- carboxysomes and metabolosomes. Although both are composed of a protein shell with encapsulated

enzymes, they differ in function, assembly and enzyme targeting. Their similarities and differences are described in the following sections.



Kerfeld CA, et al. 2010.
Annu. Rev. Microbiol. 64:391–408

Figure 1.1. Transmission electron micrographs of BMCs
a) The cyanobacterium *Synechocystis* PCC6803 with a single carboxysome (P. Shih and C.A. Kerfeld, unpublished data). (b) Pdu and (c) Eut BMCs in *Salmonella* (courtesy of J. Shively and the late H. Aldrich). (d) BMCs in *Clostridium kluyveri* grown on ethanol and acetate (courtesy of R. Lurz). (e) BMCs in *Clostridium phytofermentans* grown on fucose (courtesy of J. Blanchard). (f) Putative BMCs in *Pirellula staleyi* (courtesy of M. Rohde). Taken from Kerfeld et al. 2010. An example of a microcompartment from a-f is highlighted with a red circle.

1.1.1 α and β–Carboxysomes

Despite the functional diversity of BMCs, they tend to be conserved in their basic architecture. The first microcompartment to be characterized, the carboxysome, is found in many chemoautotrophic bacteria and all cyanobacteria (Kerfeld and Melnicki, 2016). It constitutes the main part of the cyanobacterial Carbon Concentrating Mechanism (CCM) (Badger and Price, 1992; Price et al., 2008) a system that involves the active uptake of inorganic carbon into the cell and its intracellular accumulation, primarily as HCO3−. HCO3− diffuses through the carboxysome shell; within the lumen, an encapsulated carbonic anhydrase (CA) converts it into CO2, promoting the carboxylation reaction of RuBisCO (Kerfeld and Melnicki, 2016).

1.1.2 Metabolosomes

Metabolosomes are a subgroup of BMCs that differ from carboxysomes in their function, encapsulating pathways for catabolic reactions, as opposed to the anabolic carbon fixation pathway found in carboxysomes. The function of these microcompartments is to sequester enzymes to enhance the enclosed metabolic pathway, and to protect the bacterial cell from the toxic and volatile intermediates produced during metabolic reactions (Brinsmade et al., 2005; Penrod and Roth, 2006; Sampson and Bobik, 2008). They tend to fall under three main BMC types; propanediol utilisation microcompartments (PDU) (Bobik et al., 1999) ethanolamine utilisation (EUT) (Kofoid et al., 1999) and ethanol utilisation (ETU) (Heldt et al., 2009). However, advances in bioinformatics techniques have allowed for the identification of many different types of BMCs among 19 out of the 29 established

bacterial phyla (Axen et al., 2014). This introduction will focus on the Pdu, Eut and Etu BMCs, describing both their metabolic pathways and their microcompartment shell formation, as well as recent advances in synthetic microcompartment formation and enzyme targeting.

1.1.2.1 Propanediol utilisation (PDU) micrcompartments

The first Pdu locus to be discovered was by Bobik et al. in 1999. They described the pdu operon of *Salmonella enterica serovar Typhimurium* LT2 as having genes associated with coenzyme B12 – dependent catabolism of 1,2 – propanediol. It was shown there were also five genes associated with carboxysome-like proteins, which make up the shells of the Pdu microcompartment, and these polyhedral-like shells were only formed during bacterial growth on propanediol (Bobik et al., 1999). Since then, Pdu BMCs have been shown in a range of different bacteria, including *Citrobacter freundii* (Parsons et al., 2008), *Clostridium phytofermentans* (Petit et al., 2013) and *Lactobacillus reuteri* (Sriramulu et al., 2008). The production of 1,2 – propanediol from the breakdown of the plant cell wall sugars fucose and rhamnose, is an ideal carbon and energy source for enteric bacteria, where access to plant matter is widely available (Kerfeld et al., 2010). In this particular pathway, encapsulation of enzymes and substrates away from the bacterial cytosol is thought to be advantageous to the cell, as the propionaldehyde intermediate is highly reactive and is thought to have an adverse effect on cell function (Bobik et al., 1999; Rondon et al., 1995).

1.1.2.2 Ethanolamine utilisation (Eut) microcompartments

Ethanolamine is an important carbon and energy source for mammals, produced from the degradation of the gastrointestinal membrane component phosphatidyl-ethanolamine (Stojiljkovic et al., 1995). In a similar way to the pdu operon, the eut operon genes encode homologues of carboxysome shell proteins (Stojiljkovic et al., 1995). The Eut microcompartments are most widely studied in both *Salmonella enterica* and *Escherichia coli* (Jakobson et al., 2017; Kofoid et al., 1999; Penrod and Roth, 2006; Pitts et al., 2012; Roof and Roth, 1992), two bacteria common to the mammalian gut. Again, one of the intermediates of this breakdown pathway, this time acetaldehyde rather than propionaldehyde, is highly reactive and therefore necessary to be contained within the microcompartment (Cheng et al., 2008; Kerfeld et al., 2010).

1.1.2.3 Ethanol utilisation (Etu) microcompartment

A study by Heldt et al. in 2009 suggested that there were microcompartment shell proteins (EtuA and EtuB) encoded on the genome of *Clostridium kluyveri*, that share 60% sequence identity with the carboxysome protein CsoS1A. This potential operon also contains genes for two ethanol dehydrogenases and three aldehyde dehydrogenases, as well as the two shell protein genes (Heldt et al., 2009). *Clostridium kluyveri* is unique among the clostridia as it grows strictly anaerobically on ethanol and acetate as energy sources (Seedorf et al., 2008), meaning that encapsulation of these important substrates could be essential to cell function.

1.2 BMC metabolic pathways

One of the main differences between carboxysomes and metabolosomes is in their carbon source breakdown pathways (Figure 1.2). Carboxysome-containing autotrophs are found in environments where dissolved inorganic carbon (DIC) levels are below the necessary threshold to support efficient RubisCO-mediated CO_2 fixation. The microcompartment works by increasing the concentration of DIC in proximity to the active site of RubisCO, which helps to overcome the poor affinity of the enzyme for its substrate. The BMC shell works to lessen the amount of DIC that can escape from the microcompartment (Kerfeld et al., 2010; Rae et al., 2013).

1.2.1 α-carboxysome and β-carboxysome metabolism

α- and β- carboxysomes are both responsible for encapsulating RuBisCO and CA, and have a similar enzyme pathway although they have two different forms of the RuBisCO molecule, type 1A (α) and 1B(β)(Rae et al., 2013). In this pathway, the carbonic anhydrase turns over bicarbonate to CO_2 in a cyclical reaction. RuBisCO then uses CO_2 to turnover Ribulose 1,5-bisphosphate into 3-phosphoglycerate, which leaves the microcompartment and enters the cytosol (Kerfeld et al., 2010) (Figure 1.2).

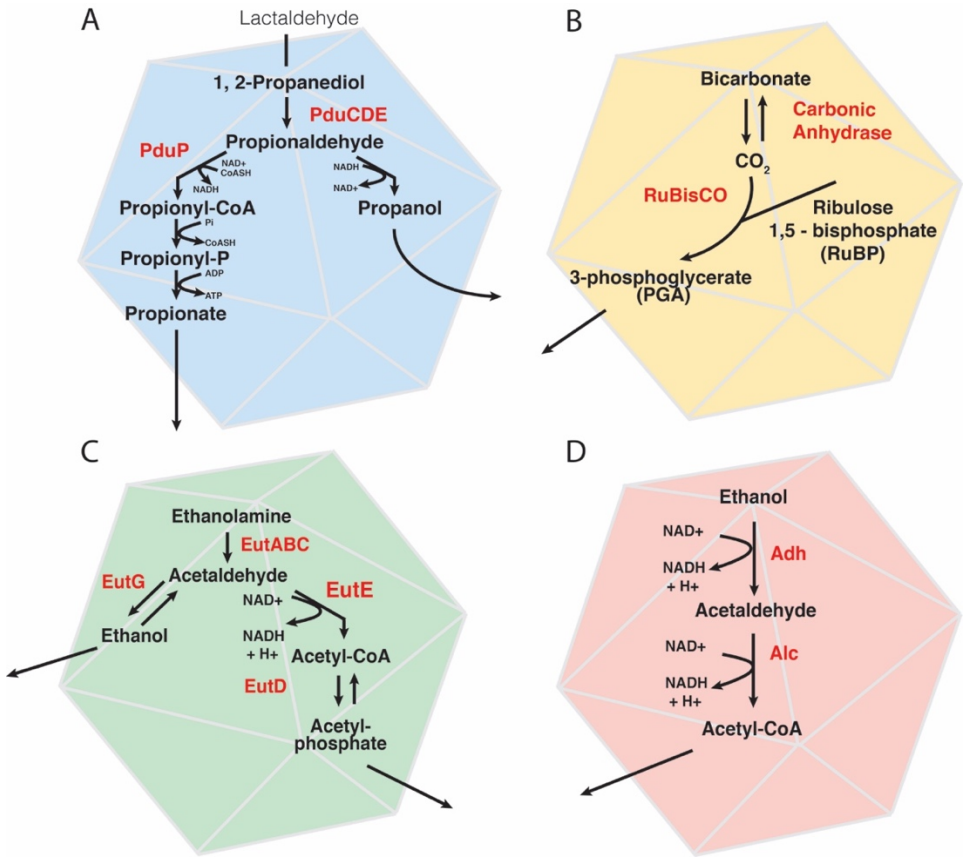


Figure 1.2. Schematic representation of simplified metabolic pathways within BMCs.
A) Fucose/rhamnose Pdu microcompartment, B) α- carboxysome, C) Eut BMC and D) Etu BMC. Figure adapted from (Kerfeld et al., 2010). Key enzymes and metabolites are highlighted for clarity, with each giving a general overview of the microcompartment enzymes required for each pathway, and show the end products that are released from the BMC back into the bacterial cell. Aldehyde dehydrogenase is abbreviated to Adh and alcohol dehydrogenase is Alc

1.2.2 1,2 – propanediol metabolism

A pathway for 1,2-PD degradation has previously been described; it is also one of the most well-characterized microcompartment pathways (Petit et al., 2013) (Figure 1.3). It starts with the conversion of 1,2-PD to propionaldehyde by coenzyme B12-dependent diol dehydratase. Then the aldehyde is converted to 1-propanol and propionic acid by propanol dehydrogenase and coenzyme A-dependent propionaldehyde dehydrogenase. A phosphotransacylase, and propionate kinase then act to produce intermediates for substrate-level phosphorylation. This process provides

ATP, reducing equivalents in the form of NADH, and potentially a three- carbon intermediate (propionyl-coenzyme A) that is degraded to pyruvate and succinate via the methylcitrate pathway. The NADH generated from this is likely to be part of a closed NAD^+/NADH cofactor pool, and explains the need for both aldehyde dehydrogenases and alcohol dehydrogenases within the BMC shell (Bobik et al., 1999; Petit et al., 2013).

Clostridium phytofermentans, a mesophilic anaerobe isolated from forest soil is another bacterium that has BMCs for breakdown of biological matter such as cellulose, into energy for the cell (Methé et al., 2002). The genome of this organism is unique in that it contains three loci encoding BMCs, one of which is for the breakdown of fucose and rhamnose (Figure 1.3). Comparative genome analysis suggests that this particular pathway may be present in other species (Petit et al., 2015).

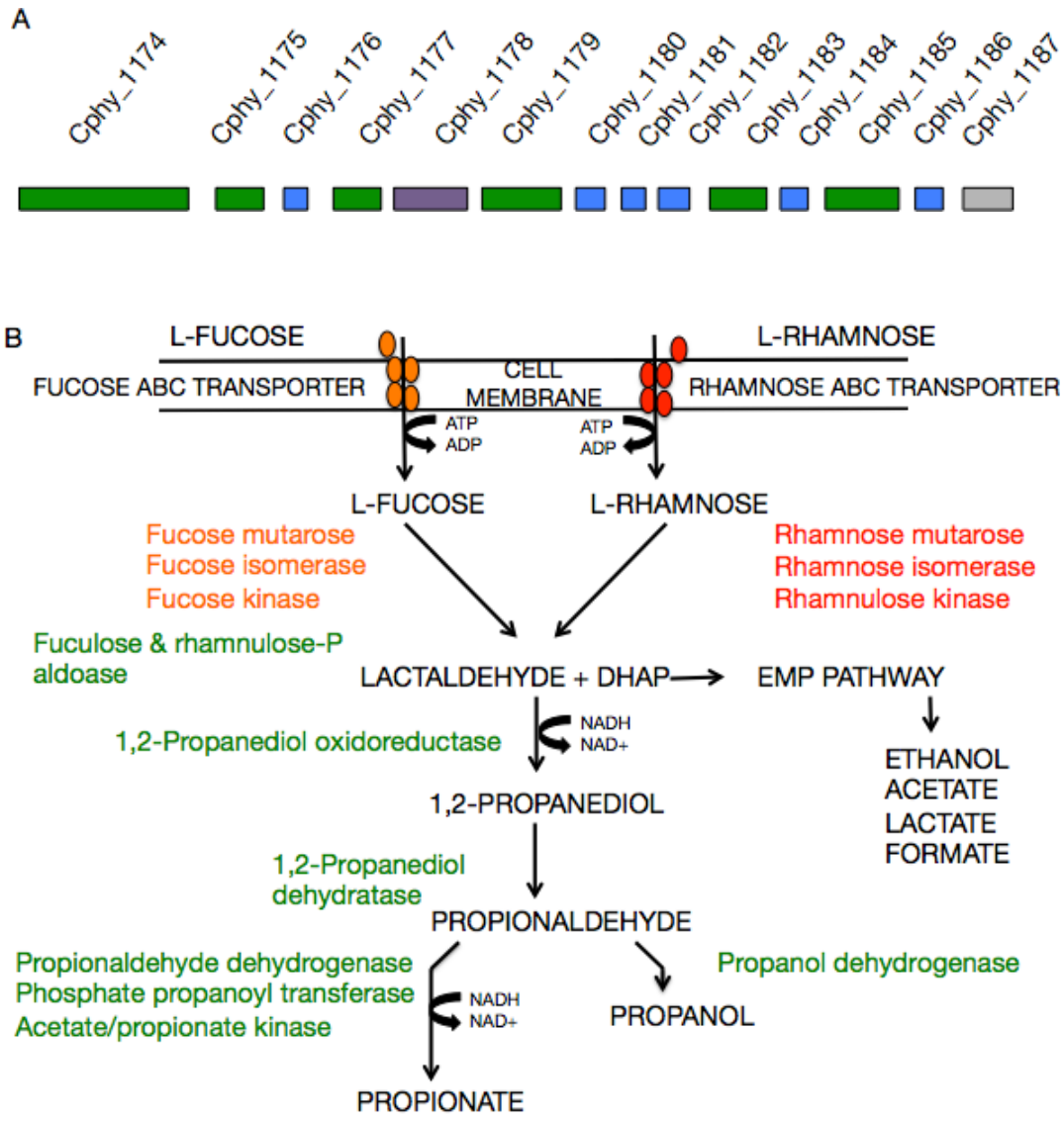


Figure 1.3. Genes associated with the fucose/rhamnose breakdown pathway in *Clostridium phytofermentans*
A) A representation of the locus of the fucose/rhamnose (Fuc) BMC. Genes involved in the pathway described below are shown in green, shell proteins shown in blue, aldehyde dehydrogenase in purple and the transcriptional regulator is shown in purple. B) A model of the C. phytofermentans fucose and rhamnose fermentation pathways, adapted from Petit et al. (2013). Enzymes shown in green are encapsulated within the BMC.

1.2.3 Ethanolamine metabolism

The pathway of ethanolamine breakdown consists of the deamination of the two-carbon compound ethanolamine into acetaldehyde, an intermediate that is thought to be toxic to the cell and therefore the primary reason for encapsulating this process

(Pitts et al., 2012). The acetaldehyde is further broken down into acetyl-phosphate, which is then utilized in substrate-level phosphorylation (Chowdhury et al., 2014). EutE is the aldehyde dehydrogenase and EutG the alcohol dehydrogenase, with the NAD^+/NADH turnover cycling between these two particular enzymes (Huseby and Roth, 2013). This cofactor pool is important biochemically, as it describes another need for the encapsulation of these enzymes (Stojiljkovic et al., 1995).

1.2.4 Ethanol metabolism

The most well studied example of the ethanol metabolism microcompartment is from *Clostridium kluyveri*. This bacterium is able to grow solely on ethanol and acetate as carbon sources, and uses an ethanol dehydrogenase to convert ethanol to acetaldehyde with NAD^+ as a cofactor. The acetaldehyde is then converted to acetyl- CoA, via an aldehyde dehydrogenase, again using NAD^+ and Co-enzyme A as cofactors. The resulting acetyl-CoA then leaves the bacterial microcompartment, and enters the Kreb's cycle, where end products butyrate and acetic acid are generated (Seedorf et al., 2008).

1.3 Glycyl-radical enzymes

Glycyl-radical enzyme (GRE) microcompartments are the most abundant type of BMC, but are only recently beginning to be experimentally characterised. The first glycyl-radical enzyme microcompartment to be characterized, was the fucose/rhamnose BMC from *Clostridium phytofermentans* (Petit et al., 2013), with the main enzyme of this pathway reported to be a 1,2-propandiol dehydratase. There are 5 distinct microcompartments within GRE microcompartments that are assumed to be

different in their function (Kerfeld et al., 2018), with the differences coming down to whether the enzyme is B12-independent or B12-dependent. The GRM5 described here is B12-independent, whereas the Pdu operon in *Salmonella enterica* relies on vitamin B12 for the catabolism of 1,2-propanediol to propionaldehyde (Bobik et al., 1999; Jeter, 1990).

To detoxify the aldehydes produced from GREs, either from B12-dependent or B12-independent pathways, an aldehyde dehydrogenase is needed to resolve the acyl-coenzyme A intermediate and produce an NADH molecule from the turnover of NAD^+ (Jorda et al., 2013).

This project will focus in detail on the biochemistry and structure of the aldehyde dehydrogenase from the microcompartment described in Petit et al. (2013) above

1.4 Overview of BMC protein shell structure

The protein shells of bacterial microcompartments are composed of three different protein domains, BMC-H, BMC-P and BMC-T. A mix of these three proteins types is able to form the canonical pseudo-icosahdral BMC shell, with the BMC-H and BMC-T proteins making the flat-faces of the icosahedron, with the vertices formed by pentameric proteins (Pitts et al., 2012). Most BMC shell proteins adopt structures in which the hexamer has a bowl-shaped inversion on one side, specifically the side on which both the N- and C- termini reside in the typical BMC domain and the BMC domain adopts an α/β-fold with a central four-stranded antiparallel β-sheet flanked by small α-helices (Kerfeld et al., 2005).

BMC-H forms a 6-fold symmetrical hexamer and is the most common type of domain in the microcompartment (Figure 1.4). BMC-T is the next most abundant type and its

monomer consists of a tandem repeat of the BMC domain, which comes together to form pseudo-hexameric trimers. Finally, the least abundant, and most structurally divergent BMC protein is the BMC-P, which forms a pentameric structure that is thought to assemble the icosahedral vertices of the microcompartment (Lassila et al., 2014) (Figure 1.5). The paper from Lassila et al. (2014), also suggests that the BMC-H protein shells are likely to only form flat sheets without the presence of the BMC-T protein, meaning that BMC-T gives curvature to the microcompartment, and allows the BMC-P protein somewhere to bind.

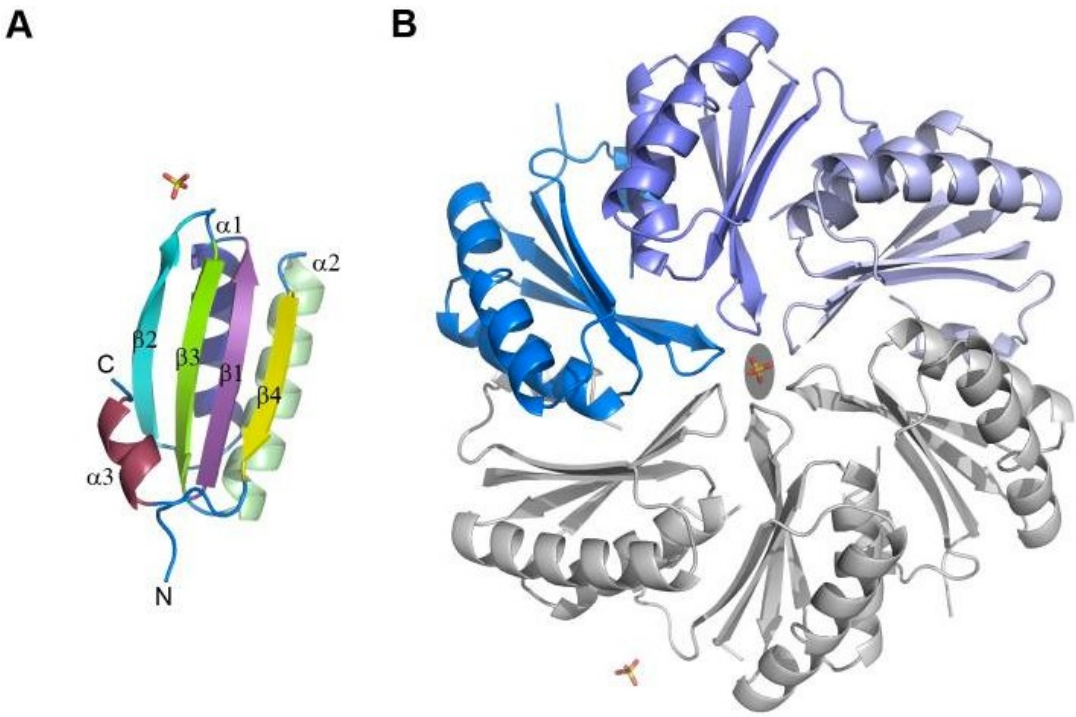


Figure 1.4. A cartoon representation of the crystal structure of the EutM shell protein from Clostridium difficile
A)A cartoon representation of the EutM monomer in the C. difficile ethanolamine utilizationmicrocompartment, with secondary structure elements coloured and labelled from N- to C- termini. B) Hexameric arrangement of EutM, showing the asymmetric unit contents in blue and symmetry related molecules in grey, with the two-fold symmetry axis generating this arrangement shown as a grey circle. Small molecules highlighted with a red circle are phosphate ions (Taken from Pitts et al. 2012)

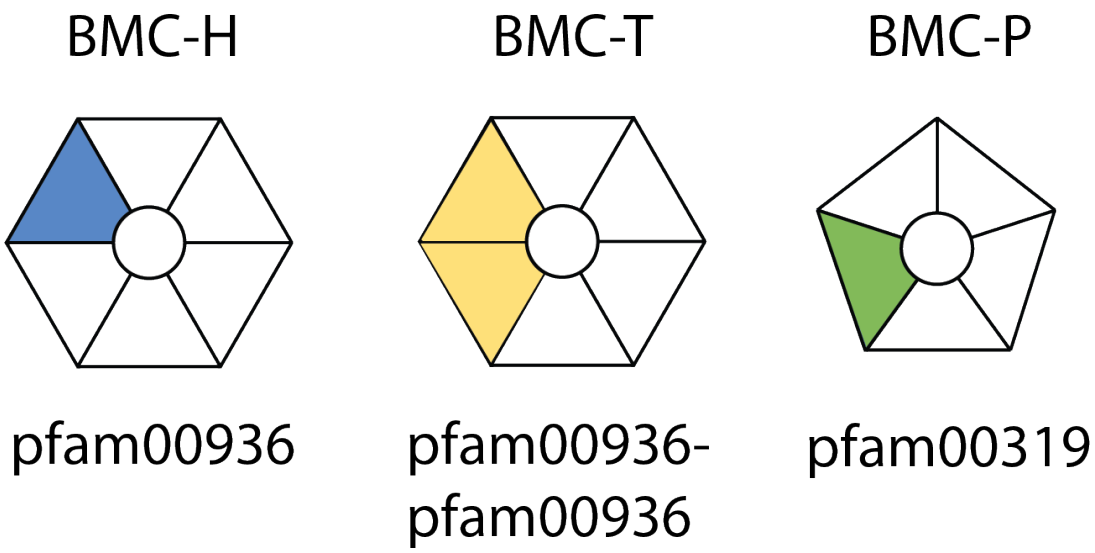


Figure 1.5. A representation of the three main types of bacterial microcompartment shell proteins and their pfam domains
BMC-H is a hexameric domain, BMC-T is a tandem hexamer domain and BMC-P is a pentameric domain. These three shell protein types make up all metabolosomes studied to date, with BMC-H being the most abundant type, followed by BMC-T and finally BMC-P. BMC- T has two subtypes, based on oligomeric status, either single or double layers (Kerfeld et al., 2018).

1.4.1 Hexameric domains (BMC-H)

The most abundant type of bacterial microcompartment shell protein is the BMC-H hexameric domain. It contains 6 single Pfam00936 domains, assembled into a cyclic homohexamer, with both a convex and concave side (Kerfeld et al., 2005). An example of a BMC-H protein from the Clostridium phytofermentans ethanolamine BMC is shown in Figure 1.4. BMC-H proteins form the faces of the bacterial microcompartment shells, with the concave side of the homohexamer facing out into the cytosol (Kerfeld et al., 2018).

1.4.2 Tandem hexameric domains (BMC-T)

BMC-T proteins are made up of a tandem fusion of two Pfam00936 domains and as such form trimers that resemble the BMC-H hexamers. The two subtypes of BMC-T are BMC-Ts (single) and BMC-Td (double), the former is a single trimer and the latter has two trimers dimerizing across their concave faces (Kerfeld et al., 2018) (Figure 1.4).

1.4.3 Pentameric domains (BMC-P)

BMC-P proteins (Pfam03319) are pentameric in shape, and form the vertices of the microcompartment shell. They are much less abundant than the BMC-H and –T proteins, however they are thought to be essential to provide the barrier for a completely closed microcompartment shell (Cai et al., 2009).

1.4.4 Microcompartment shell assembly

Despite α- and β- carboxysomes being similar in terms of their enzyme pathways, with both utilizing RuBisCO and CA for CO_2 fixation, it has been shown that the β-carboxysomes are much more closely related to metabolosomes than they are to α-carboxysomes (Cameron et al., 2013). This is apparent most in the assembly of the β-carboxysome and metabolosome microcompartment shells, with the β-carboxysome core assembling first followed by encapsulation of the shell. As metabolosome core enzymes tend to aggregate and have putative localization sequences that are thought to interact with the shell (Fan et al., 2010), it is likely that their encapsulation pathway is also core-first. However, in α-carboxysomes the protein shell and core proteins are

likely to be assembled at the same time (Iancu et al., 2010). In the context of synthetic biology, the difference in assembly methods between carboxysome and metabolosome types could have a big impact in terms of building synthetic microcompartment systems, particularly with regards to targeting enzymes to the microcompartment shell.

1.5 Bacterial microcompartment enzyme-shell targeting

Although it is likely that there is a difference how key pathway enzymes are targeted to the microcompartment, one particular study has shown that an N-terminal signal peptide is vital for the encapsulation of the PduP (propionaldehyde dehydrogenase) from the Pdu microcompartment in *Salmonella enterica* (Fan et al., 2010) with another study showing the same effect in *Citrobacter freundii* (Lawrence et al., 2014). These enzymes were localized via a short, N-terminal aldehyde dehydrogenase localization sequence, 18 amino acids in length. In the paper by Fan et al., (2010), they identify the targeting site for the enzyme localization sequence, which in this instance is the H81, V84 and L88 of the PduA shell protein C-terminus.

Many studies so far have looked at bacterial microcompartments with a view to engineering them to make synthetic compartments that can be utilized in biofuel production, including creating empty carboxysomes (Menon et al., 2008) overexpression of BMC shell proteins from a Pdu microcompartment (Parsons et al., 2008), formation of nanotubes using a BMC-H protein (Noël et al., 2016) and engineering of BMC shell proteins to introduce a new function (Aussignargues et al., 2016). However, these tend to use classical cloning methods, which are costly and time consuming, particularly if large numbers of constructs need to be produced. Advances

in synthetic biology cloning techniques mean that BMCs have the potential to become extremely useful tools for a wide range of applications.

1.6 BMCs as tools for synthetic biology

Interest in the study of BMCs is growing as research moves towards synthetic biology and the use of standardized parts as platforms for biotechnology and metabolic engineering. By studying how the microcompartment shells are formed and how the enzymes used in the breakdown pathways are recruited to the BMC, it should be possible to synthesize empty microcompartments, into which biologically useful enzyme pathways can be added. These synthetic microcompartments could then be used to optimize efficiency of pathways used in the synthesis of fine drugs and chemicals, as well as to help reduce toxicity and the release of volatile intermediates into host cells. The examples in the following sections highlight the cutting-edge work being carried out into microcompartment bioengineering.

1.6.1 Bacterial microcompartments in bioengineering

An example of using bioengineering to enhance the modularity of BMCs was shown by Parsons et al. (2008). This study took the 21 gene Pdu operon from *Citrobacter freundii* and showed that it was fully functional within *E. coli* cells, for both microcompartment formation and metabolic activity (Parsons et al., 2008). Carboxysomes have also been highly engineered, with a study showing that the alpha-carboxysome of the chemoautotroph *Halothiobacillus neapolitanus* has been expressed in *E. coli*, with the resulting particles showing CO_2 fixing activity (Bonacci et al., 2012). Another interesting study, by (Quin et al., 2016), shows that it is possible

to take the localization sequences from two different Eut microcompartment enzymes, EutC and EutE and use them to target GFP to a specific microcompartment shell protein, EutS, which ultimately allows them to be encapsulated. These relatively recent examples show that all microcompartments are potentially heterologous, which has interesting implications for future rational design of synthetic microcompartments

1.7 Golden Gate cloning and Type II restriction enzymes

To utilize the inherent modular properties of BMCs and make them into a useful toolkit system would be expensive and time-consuming using standard cloning methods. The recent advances in Golden-Gate cloning, using Type II restriction endonucleases means that it is possible to build a system of BMC 'parts' quickly and effectively.

Techniques such as 'gap repair cloning' were first used as an approach to synthetic biology and high-throughput cloning. In a yeast-cloning system, this involves amplifying the ORF with ORF-specific primers that have flanking sequences homologous to the ends of a linearized cloning vector (Marsischky and LaBaer, 2004). A similar system was developed for *E. coli* as prokaryotic system, using a λ Red/ recET- mediated recombination (Datsenko and Wanner, 2000; Marsischky and LaBaer, 2004). Downsides to this cloning method include the fact that it is not possible to easily transfer clones between different vectors and the PCR primers necessary for this technique are fairly long and therefore prone to synthesis errors (Marsischky and LaBaer, 2004). This is an example of one of many techniques that has tried to improve upon standard cloning methods, without satisfying all of the needs of fully robust system. Ideally, synthetic biological parts would be easy to transfer between different vector systems, quick to clone (< one day), and inexpensive to make, with a scar-less

insertion into the acceptor vector. The next section describes some recent exciting techniques to improve upon current bioengineering efforts.

1.7.1 Type II restriction endonucleases

Type II restriction endonucleases such as BsaI work on the principal of DNA cleavage outside of their recognition site, resulting in 5' or 3' DNA overhangs that can be a specified nucleotide sequence. This gives 256 possible different overhang sequences that can be designed with multiple fragment assembly in mind (Engler et al., 2008). Two defined DNA ends can be designed to be flanked by a type II restriction site, so that the enzyme recognition sequence is removed, but a user-specified overhang sequence remains (Figure 1.6). This property allows cloning to be performed using a one-step restriction-ligation. This strategy was shown to result in the conversion of more than half of all input plasmids present into the desired recombinant product in just a 30 minute restriction-ligation reaction (Engler et al., 2008).

1.7.2 CIDAR MoClo: Modular assembly standard

The strategy to assemble synthetic BMCs needs to be time and cost-effective, as well as being able to generate large DNA libraries in a rational way. The Modular cloning, or MoClo system (Iverson et al., 2015; Weber et al., 2011) is a multipart, one-pot digestion system that allows users to generate specific overhangs for each 'part', such as the RBS or a CDS and insert them in to a specific acceptor vector system. As the system uses type II restriction endonucleases, once the enzyme has cut outside of the recognition site, the recognition site disappears. This allows subsequent cycles of digestion and ligation for maximal plasmid insertion. The MoClo system works on a

plasmid ‘hierarchy’ – the Level 0 plasmids are used to put user-defined ‘parts’ into the system. The next level builds the Level 0 parts into transcriptional units, using Level 1 vectors with kanamycin resistance genes, rather than ampicillin (Level 0 and Level 2). The differences in antibiotic resistance between the levels mean that it is easy to establish which constructs have been produced. The Level 2 plasmids join together two separate transcriptional units, which goes some way to alleviating the problem that a maximum of only six parts can be cloned into a Level 1 vector.

The MoClo system has a library of often-used genetic parts, including promoters, RBSs, reporter proteins (GFP, YFP), and terminators, as well as a 3 part vector system to insert DNA parts in to build transcriptional units (Figure 1.6 and Table 1.1). The basic parts to build the toolkit were taken from either the BioBrick registry (http://partsregistry.org), or Addgene, a useful parts repository. The BioBrick registry is developed as part of the International Genetically Engineered Machine (iGEM) competition, that involves students worldwide building biologically useful standardised parts, as part of the competition. This registry is a valuable source of Anderson promoters, reporter proteins and chromoproteins, along with plant and Bacillus subtilis chassis and parts for metal sensing and binding (Casini et al., 2014; Osbourn et al., 2012; Sarrion-Perdigones et al., 2011).

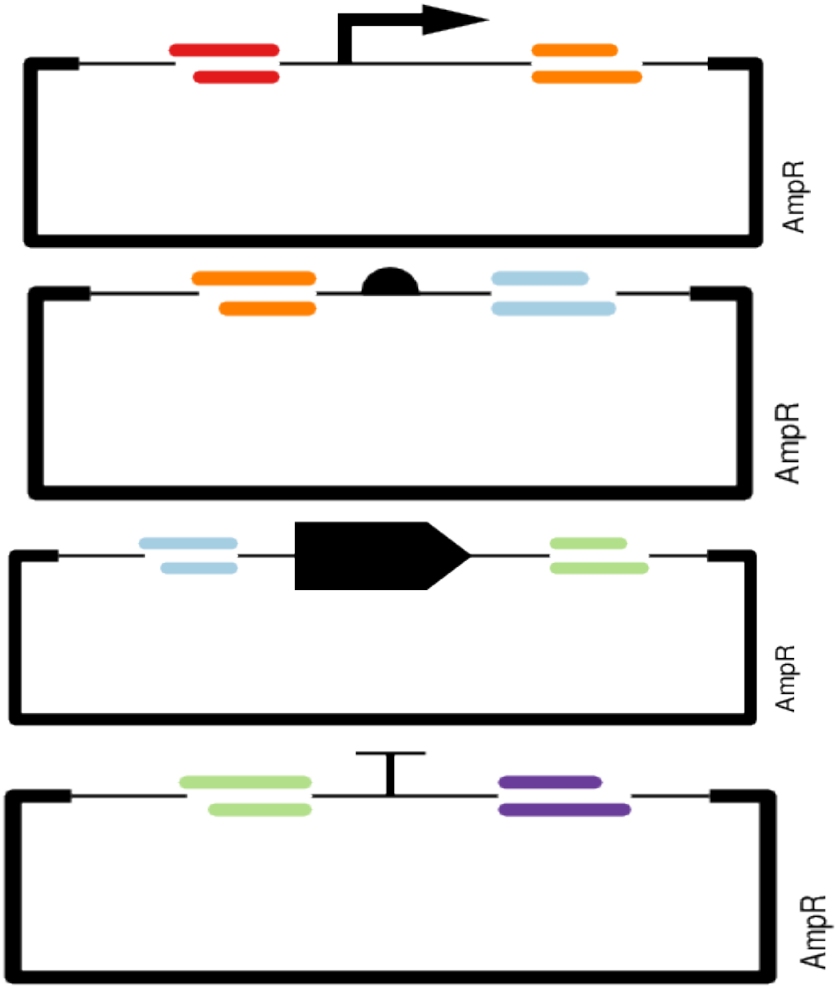


Figure 1.6 SBOL generated figures to show the assembly of the MoClo transcriptional units. Colours refer to a MoClo specified sequence, shown in Table 1.1 with the promoter (top left), RBS (top right), ORF (bottom left) and terminator (bottom right) Level 0 plasmid being mixed together in a one-pot assembly, using a Type II restriction endonuclease and BsaI ligase to cut and insert the fragments into a transcriptional vector (Iverson et al., 2015; Werner et al., 2012).

Part type	5' overhang	3' overhang
Promoter	A - GGAG	B - AGTA
RBS	B - TACT	C - CATT
Reporter gene	C - AATG	D - ACCT
Terminator	D - AGGT	E - AAGC
Acceptor plasmid	A - CCTC	E - GCTT

Table 1.1 5' and 3' toolkit defined overhangs of the MoClo system
Colours of overhangs co-ordinate to SBOL images in Figure 1.6.

1.8 Aims and objectives

- Chapters 2 and 3: The first aim is to characterise the aldehyde dehydrogenase from the Clostridium phytofermentans fucose/rhamnose Pdu microcompartment. This thesis will describe the Michaelis-Menten kinetics of the enzyme and find its biologically relevant aldehyde substrate, using NAD+ turnover assays. Control active site variants will also be assayed. As well as this, the first crystallographic structure of an acylating aldehyde dehydrogenase with CoA in the cofactor binding site will be shown, with this data helping to propose a mechanism of action for the aldehyde dehydrogenase. Knowledge gained from this work will help further the understanding of a complicated enzyme mechanism, for future studies into overall BMC biology.
- Chapter 4: As part of the CIDAR MoClo toolkit, a wide range of constitutive and controllable promoters, high, medium and low strength RBSs, a number of fluorescent reporter genes and terminators with suitable overhangs for different levels of the system are available. The toolkit parts will be used to construct a shell protein library, to develop various transcriptional units with different numbers of shell proteins from different bacterial species, with a view to expressing and visualizing these synthetic microcompartment shells with TEM. The second objective is to design RBS and localisation sequence constructs with the localisation sequences from Salmonella and C. phytofermentans Pdu microcompartments. The rationale behind this is to find out if a fixed protein part downstream of the different strength RBSs will lead to differing expression levels of fluorescent reporter, or if there is a particular strength RBS that will be optimal for production of the reporter. This can hopefully be used to test targeting a specific amount of protein to a microcompartment. The final objective

is to look at EutM shell proteins across different species with our collaborators in the Schmidt-Dannert lab. The EutM proteins have been shown to have differing charge states across their surfaces, particularly around the pores that are responsible for the influx and efflux of metabolites from BMCs. It would be interesting to get a three-dimensional view of the EutM shell protein, either by crystallising the protein or looking at protein overexpression with thin-section electron microscopy. This data would help elucidate the role of BMC shell proteins as possible scaffolds for novel enzyme pathway targeting.

- Chapter 5: This chapter is a discussion of the results from chapters 2-4.

2 Expression and purification of acylating aldehyde dehydrogenase enzymes from *Clostridium phytofermentans* bacterial microcompartment operon

2.1 Introduction

In the C. phytofermentans genome, there are three putative aldehyde dehydrogenase enzymes associated with Bacterial Microcompartment loci: Cphy1178, Cphy1428 and Cphy2642 (Petit et al., 2013, 2015). The aldehyde dehydrogenase family is split into acylating and non-acylating variants. The mechanism of the acylating aldehyde dehydrogenases is as follows:

$$\text{Aldehyde} + NAD^+ + \text{Coenzyme A} \leftrightarrow \text{Acyl-CoA} + NADH + H^+$$

Aldehyde dehydrogenases play an important role in the sugar breakdown pathways of many bacterial microcompartments, mentioned in detail in Chapter 1. The homologue of the putative enzymes mentioned above is the probable aldehyde dehydrogenase from *Listeria monocytogenes* (PDBID: 3K9D), although the mechanism of action for this enzyme is not shown. Two ligands are annotated in the crystal structure, a chloride ion and a glycerol molecule, however neither the NAD+ or CoA cofactors are shown. The aim of this section is to recombinantly express and purify soluble Cphy1178, Cphy1428 and Cphy2642 proteins, to allow further enzyme characterisation.

2.1.1 Overexpression of recombinant Cphy1178, Cphy1428 and Cphy2642

Work to previously characterise Cphy1178, Cphy1428 and Cphy2642 was carried out by Thiau Fu Ang (Undergraduate honours student, Marles-Wright lab). The full-length, wild-type forms of these enzymes were found to be unstable, so a truncated version of Cphy1178, Cphy1428 and Cphy2642 was designed, without the proposed

N-terminal tag (shown in the protein sequence alignment in Figure 3.1). A small scale-expression test was carried out with the truncated proteins, and they were found to show a suitable amount of protein for large-scale purification. *E. coli* cells, transformed with plasmids containing genes for Cphy1178$_{(20\text{-}462)}$, Cphy1428 and Cphy2642, were cultured as described in Materials and Methods, then lysed with SDS-loading buffer and subjected to 12% w/v SDS-PAGE. The nomenclature of the truncated versions of the proteins are as follows: Cphy1178_20/Cphy1178$_{(20\text{-}462)}$, Cphy1428, Cphy2642.

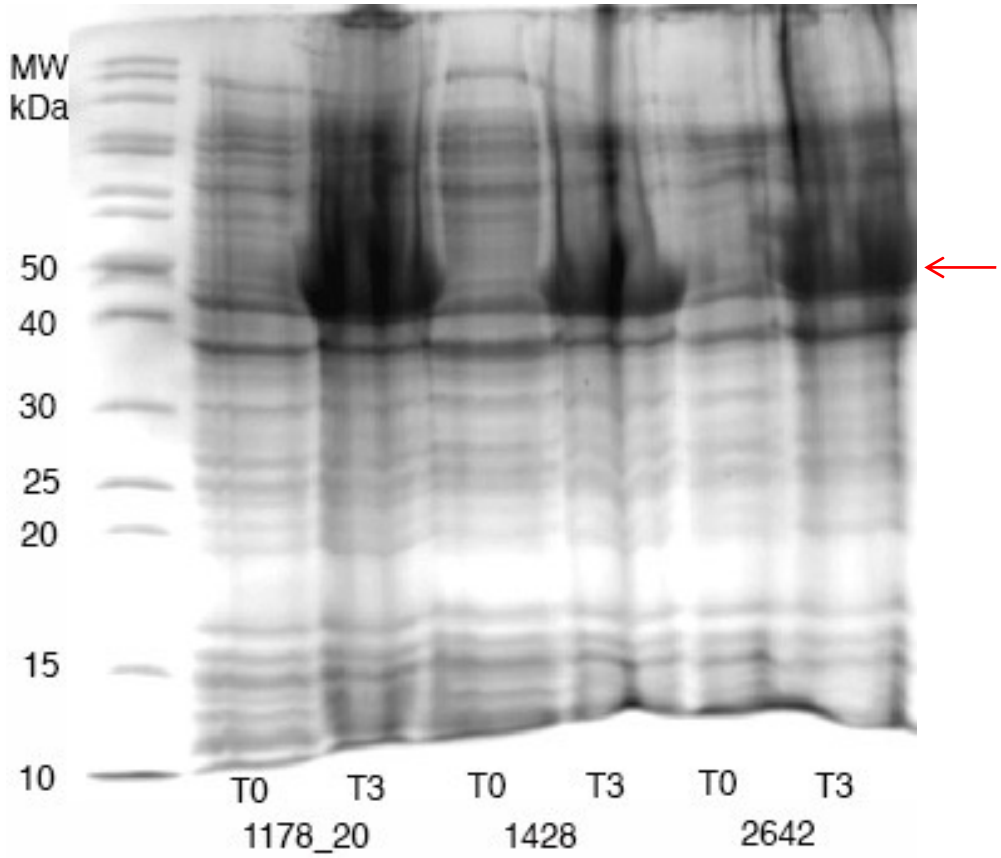


Figure 2.2 Expression tests of BMC-associated aldehyde dehydrogenases from C. phytofermentans produced recombinantly in *E. coli* BL21 (DE3) cells.
Cphy1178_20, Cphy1428 and Cphy2642. T0 and T3 refer to 1 ml samples taken at the point of induction and three hours after induction with 1mM IPTG, respectively; then analysed by 12% SDS-PAGE. The red arrow indicates where the bands for the required enzymes are expected based up the calculated molecular weight (approx. 50 kDa for all three proteins).

Protein	Molecular weight (kDa)	Theoretical pI	Extinction co-efficient (M^{-1} cm^{-1})
Cphy1178_20	47.5	6.41	16,960
Cphy1428	52.0	6.08	18,910
Cphy2642	50.1	8.22	14,440

Table 2.1 Calculated parameters of the three aldehyde dehydrogenase enzymes
Parameters were calculated using Kyoto Encyclopaedia of Genes and Genomes (KEGG) for the protein sequence, and ProtParam (ExPASy) software.

The protein expression tests indicated that all three proteins were overexpressed to high levels and were therefore amenable to large-scale (1 litre) expression and purification. However, as an exploratory expression test, the proteins were not tested for solubility at this stage, just overall expression in the cells.

2.1.2 Purification of recombinant aldehyde dehydrogenases by anion exchange and size-exclusion chromatography

Once the appropriate expression conditions of induction had been established (1 mM IPTG after cell growth to an OD600 0.6) for Cphy1178_20, Cphy1428 and Cphy2642, the enzymes were purified from 1 litre of LB culture of *E. coli*. The cells were harvested, lysed by sonication and the soluble fraction clarified by centrifugation followed by filtration.

The proteins were subject to anion exchange using a Q-sepharose column (full protocol and buffer details in Methods chapter), with the chromatograms and SDS-PAGE gels shown in Figures 2.3, 2.5 and 2.7 for Cphy1178, Cphy1428 and Cphy2642 respectively. A further purification step was necessary to get a purer protein product. Samples taken and pooled from the Q-sepharose column were concentrated for further purification on a size-exclusion chromatography column. Work undertaken by Didi He in the Marles-Wright lab shows the calibration curve for the HiLoad 16/600

Superdex 200pg Size Exclusion Chromatography column (see Appendix). Pooled samples were concentrated to 2 ml and injected on to a HiLoad 16/600 Superdex 200pg Size Exclusion Chromatography column. This stage of purification was successful for constructs Cphy1178_20 and Cphy1428 (Figures 2.4 and 2.6), so fractions of the correct size were pooled and their concentration measured by nanodrop, using the MW and extinction coefficient values from Table 2.1. However no protein was observed after Size Exclusion Chromatography of Cphy2642. Protein purified from size-exclusion purification was deemed to be clean enough to use for the aldehyde dehydrogenase NAD^+ turnover assays described later in this chapter.

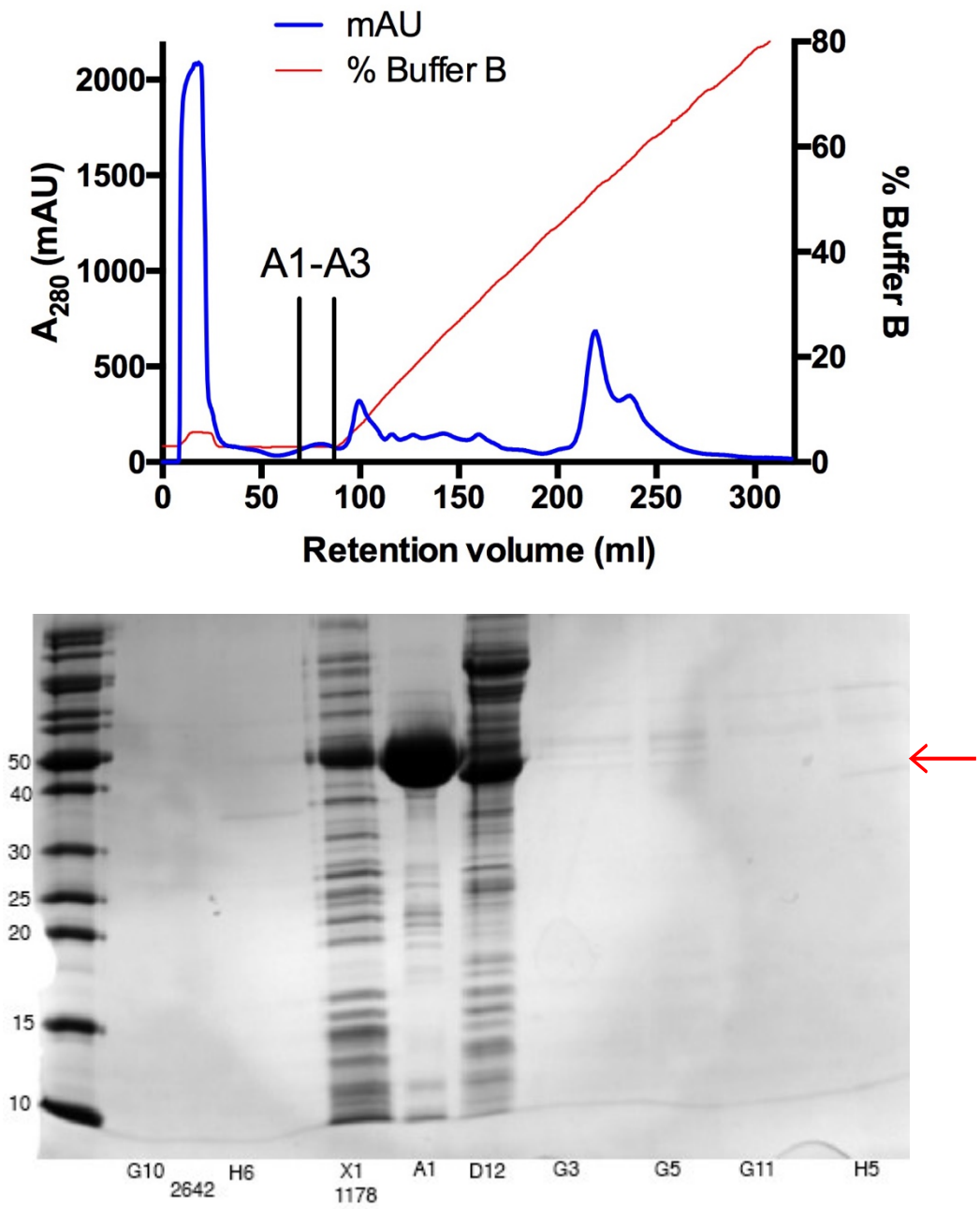


Figure 2.3 Anion exchange purification of recombinant Cphy1178.
Purification from a 12 ml Q-sepharose column. X1 corresponds to the flow-through, which contained a small percentage of unbound protein. Bound proteins were eluted by a linear gradient over 20 column volumes from 0–1 M NaCl as fractions from A1 – H5. Fractions from A) were analysed by 15% SDS-PAGE, in which fractions A1 – D12 showed a band of the appropriate size (red arrow). Fractions were pooled from A1-A3 for further analysis

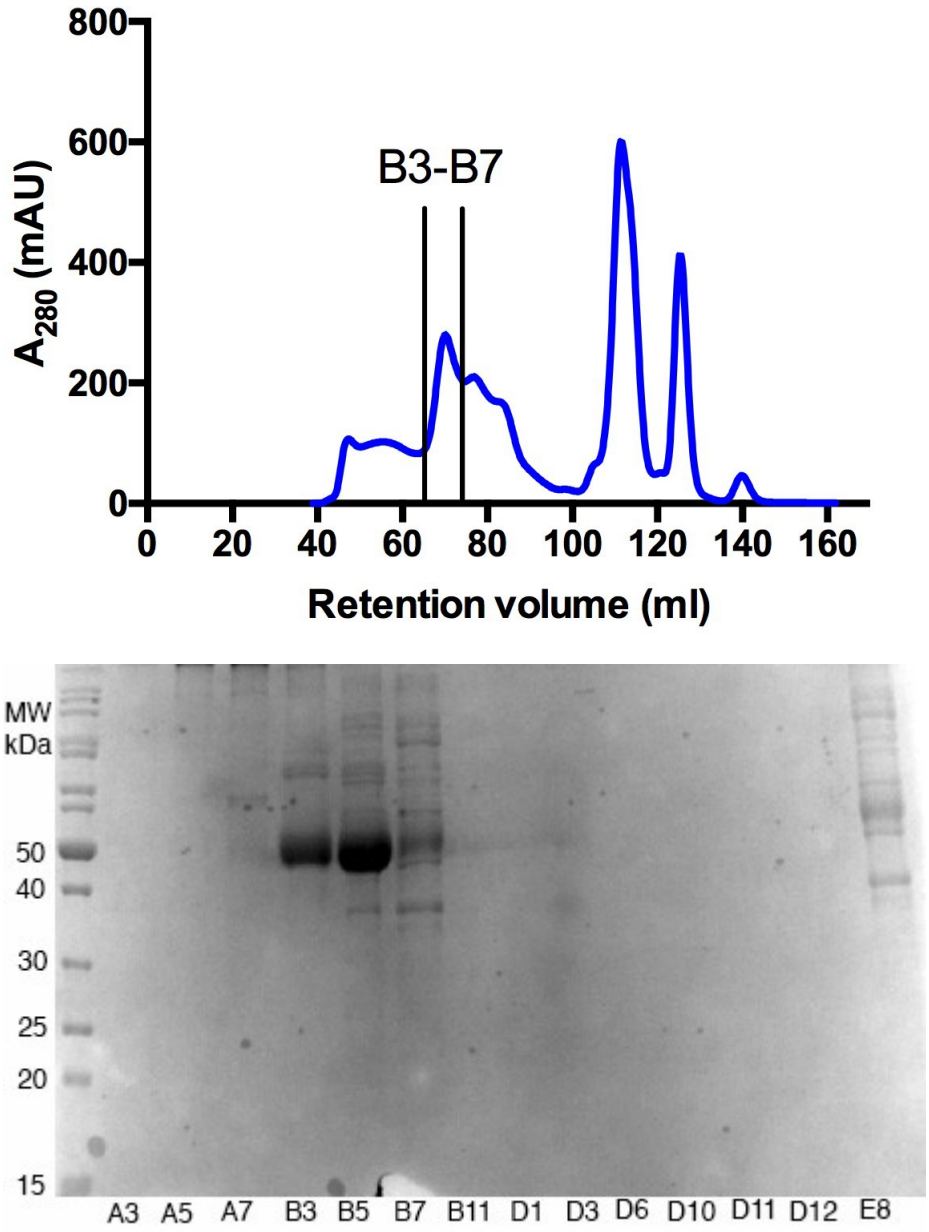


Figure 2.4 Size-exclusion chromatographypurification of recombinant Cphy1178.
Purification from a 120 ml HiLoad 16/600 Superdex column. Fractions A3 – E8 were analysed by 15% w/v SDS-PAGE gel (B). Protein of the correct size was observed in fractions B3-B5.

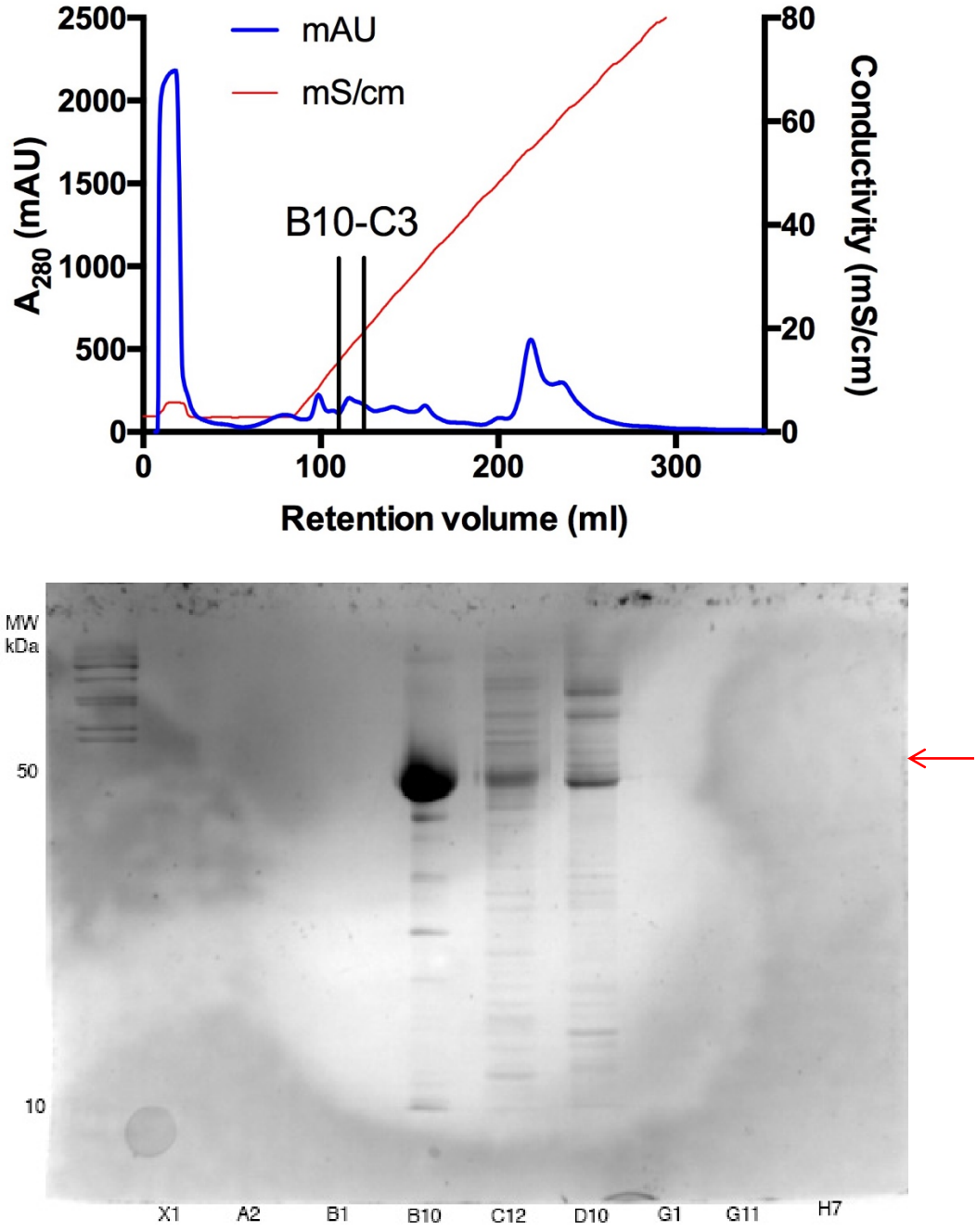


Figure 2.5 Anion exchange purification of recombinant Cphy1428.
Purification from a 12 ml Q-sepharose column. Fraction X1 corresponds to the flowthrough. Fractions were eluted by a linear gradient over 20 column volumes from 0–1 M NaCl as fractions from A2 – H7. Fractions from A) were analysed by 15% SDS-PAGE, in which only fractions B10-D10 showed a band of the appropriate size (red arrow). Fractions were pooled from B10-B12 for further analysis.

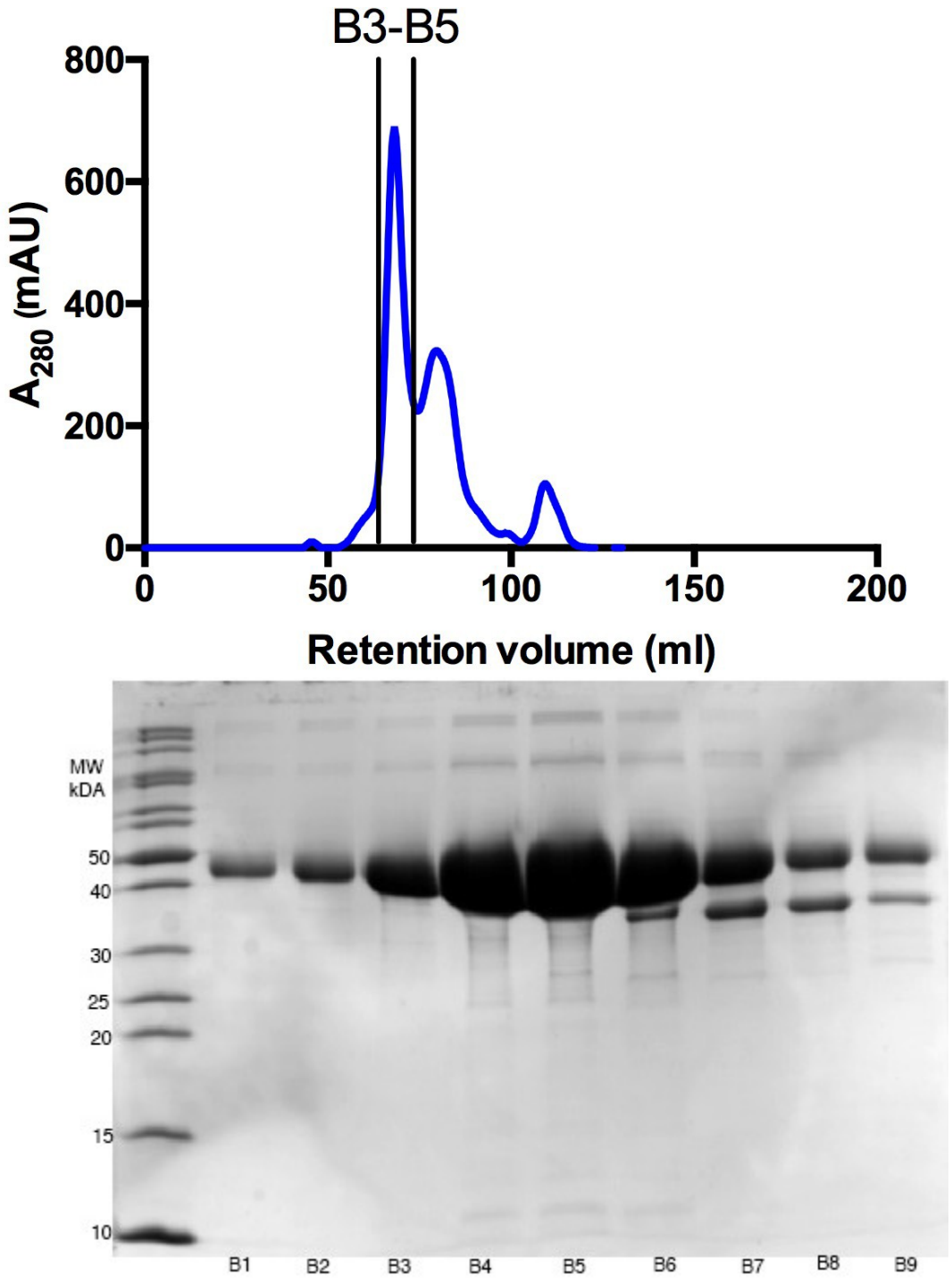


Figure 2.6 Size-exclusion chromatography purification of recombinant Cphy1428.
Purification from a 120 ml HiLoad 16/600 Superdex column. Fractions A3 – E8 were analysed by 15% SDS-PAGE gel (B). Protein of the correct size was observed in fractions B3-B5.

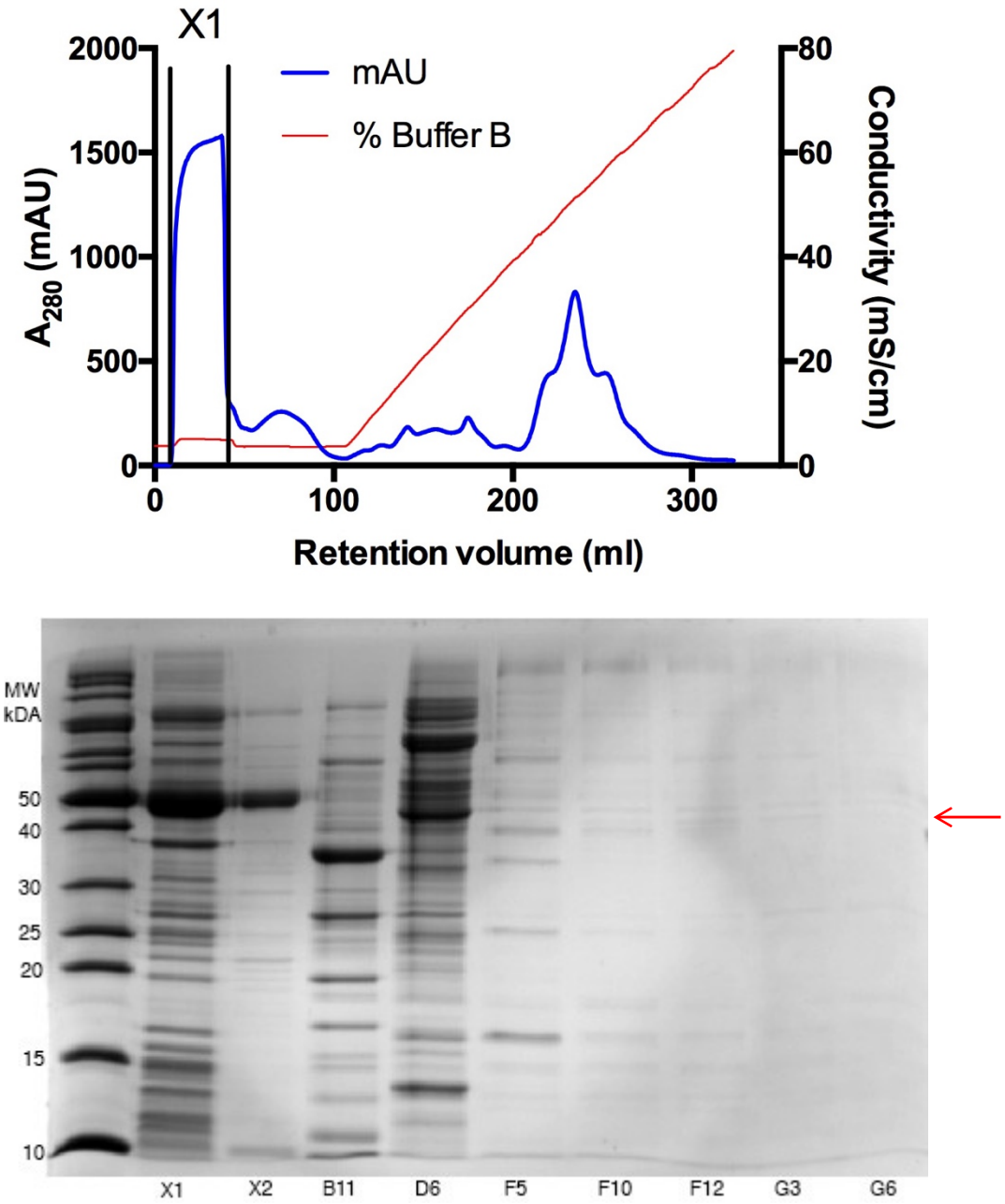


Figure 2.7 Anion exchange purification of recombinant Cphy2642.
Purification from a 12 ml Q-sepharose column. Fraction X1 corresponds to the flowthrough. Fractions were eluted by a linear gradient over 20 column volumes from 0–1 M NaCl as fractions from A1 – H7. Fractions from A) were analysed by 15% SDS-PAGE, in which only fractions X1 showed a band of the appropriate size (red arrow). Flowthrough X1 was used for further analysis.

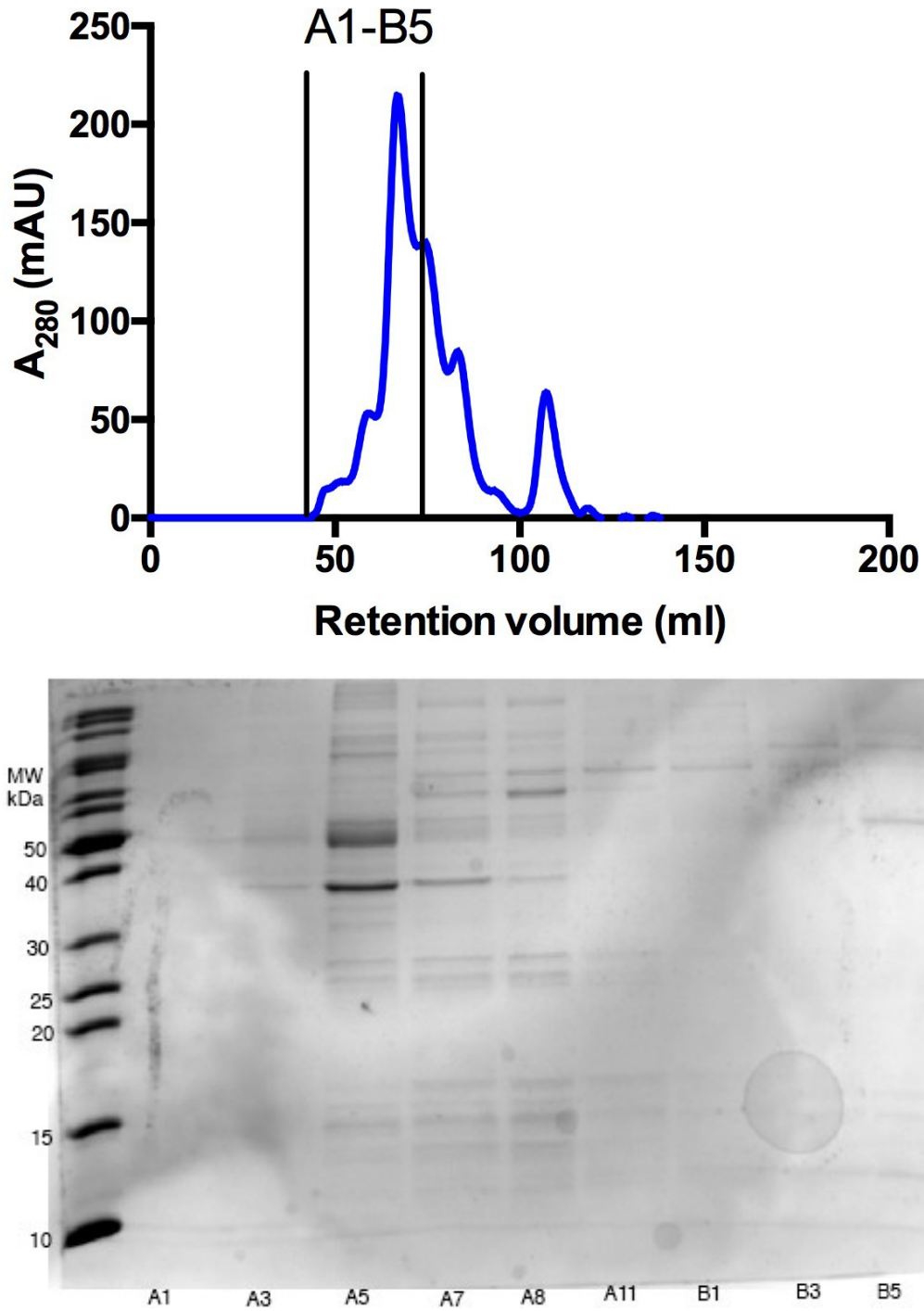


Figure 2.8 Size-exclusion chromatographypurification of recombinant Cphy2642.
Purification from a 120 ml HiLoad 16/600 Superdex column. Fractions A1 – B5 were analysed by 15% SDS-PAGE gel (B). A faint band was observed in

fraction A5, however this amount of protein was considered too low to use in further experiments.

2.1.3 Summary and Discussion

The three putative aldehyde dehydrogenase enzymes from Clostridium phytofermentans Cphy1178, Cphy1428 and Cphy2642 were expressed as recombinant proteins in E. coli BL21 DE3 cells. Initial expression tests showed that the proteins expressed well under standard conditions (37 ºC for 3 hours with shaking at 250 rpm). The enzymes were purified in two stages, first by ion-exchange chromatography, over a 12 ml Q-sepharose column and further by size-exclusion chromatography, HiLoad 16/600 Superdex column.

The enzymes were then used for future experiments, including the aldehyde dehydrogenase NAD+ turnover assays and crystal trials. Despite the initial insolubility of the full-length protein, the truncated variants stayed soluble at room temperature, and so all subsequent steps were carried out at approximately 18 ºC.

2.2 Enzyme activity of acylating aldehyde dehydrogenase enzymes from Clostridium phytofermentans bacterial microcompartment operons

2.2.1 Introduction

The enzyme activity of Cphy_1178$_{20\text{-}462}$ had been previously characterized by Thiau Fu Ang (Undergraduate honours student, Marles-Wright lab), however there is no biochemical data for other BMC associated aldehyde dehydrogenase enzymes annotated in the *C. phytofermentans* genome: Cphy2642 and Cphy1428. It would interesting to find out if these enzymes have the same or different substrate

specificities as Cphy117820-462, given the fact that they are required for distinct pathways and are likely to have different substrates.

2.2.2 Preliminary NAD+ turnover assay set-up

The characterization of these enzymes was performed, with the work on Cphy_117820-462 being repeated to try to optimize experimental conditions for the other two enzymes. As well as using 2-mercaptoethanol (BME) as a reducing agent to ensure the active-site cysteine was in an appropriately reduced state for turnover, Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was also used as a stronger reducing agent . However, no data could be generated from these two enzymes despite attempts to optimise the experimental set up. It could be the case that acetaldehyde, propionaldehyde or butyraldehyde are not the substrates of these enzymes, however this is unlikely as inspection of the C. phytofermentans genome suggest that the substrates for these enzymes are likely to be intermediates from the breakdown of ethanolamine, or choline, which are the likely functions of these BMCs (Petit et al., 2013).

2.2.3 Enzyme assay set-up

For purified proteins with activity from small-scale assays, enzymes assays were carried out to determine kinetic parameters using a 96-well micro-titre plate reader, with each well containing 300 µl of reaction buffer. All wells were made up to 300 µl volume with 100 mM KCl, 100 mM Tris pH8.0, 0.66 mM β-NAD, and Cphy1178 at varying concentrations. For the acetaldehyde assay, 250 nM of Cphy1178, A340 was measured every 7 s for 182 s and the assay was carried out in 5 replicates. However,

subsequent iterations of the experiment used 100 nM of enzyme, as this gave the best overall results in terms of catching the initial turnover step. The assay blank used standard assay components without protein. Initial rates from linear regression for NADH produced for different substrate concentrations were obtained from data analysis. These initial rates produced a curve fitted with non-linear regression applying the Michaelis-Menten equation using Prism 6 software (Graphpad), shown in Table 2.1 and Figure 2.9.

2.2.4 Enzyme assay results

By monitoring the hydride transfer step of the aldehyde dehydrogenase enzyme reaction by spectrophotometric measurement of the production of NADH in the presence of aldehyde substrates, we were only able to detect NAD+ dependent activity for Cphy1178. The other enzymes displayed no distinguishable activity with either NAD+, or NADP+, for the substrates used in this study. This may be due to issues with the instability of these proteins out of their native environment within the BMC, or a requirement for some additional factor not present in our assays.

We were however able to detect activity against short-chain fatty aldehydes with up to seven carbon atoms (Figure 2.9) for Cphy1178, although longer chain aldehydes were difficult to assay reliably due to their insolubility in biological buffers and we were unable to determine accurate kinetic parameters for C7 and higher aldehydes (Table 2). Cphy1178 enzyme displayed the lowest KM and the highest kcat/KM values for the substrates tested. This data is consistent with the hypothesis that this protein acts as a propionaldehyde dehydrogenase within the C. phytofermentans fucose/rhamnose BMC.

It is noteworthy that this enzyme displays substrate inhibition at high concentrations of aldehydes, shown by a decrease in the velocity values in the graphs in Figure 2.9; interestingly this is more marked with aldehydes with an odd number of carbon atoms. This substrate inhibition is consistent with previous reports on the activity of yeast ALDH2 enzyme (Wang et al., 2009). The reason they suggest for this is that at higher substrate concentrations, a dead-end enzyme- NADH-aldehyde complex may be formed.

Substrate	K_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (s^{-1} mM^{-1})	K_i (mM)
Acetaldehyde	1.62 ± 0.02	5.16 ± 0.10	0.31	n/a
Propionaldehyde	3.45 ± 0.03	0.82 ± 0.02	4.21	17.31 ± 0.46
Butyraldehyde	3.44 ± 0.7	1.74 ± 0.07	1.98	144.3 ± 30.2
Pentanaldehyde	5.44 ± 0.10	2.2 ± 0.7	2.47	12.61 ± 0.46
Hexanaldehyde	5.61 ± 0.12	2.90 ± 0.10	1.93	23.52 ± 1.28

Figure 2.9 Activity of Cphy1178 against various aldehyde substrates.
Cphy1178 was incubated with increasing concentrations of aldehyde substrates in the following reaction mixture: 100 mM Tris.HCl (pH 8.0); 0.66 mM NAD+, 100 mM KCl, 10 mM 2- mercaptoethanol. Data were analysed using Graphpad Prism and kinetic parameters were calculated by non-linear regression to the Michaelis-Menten equation with substrate inhibition for C3 and larger aldehydes. (Velocity is expressed as mM NADH s-1 mM-1ADH).

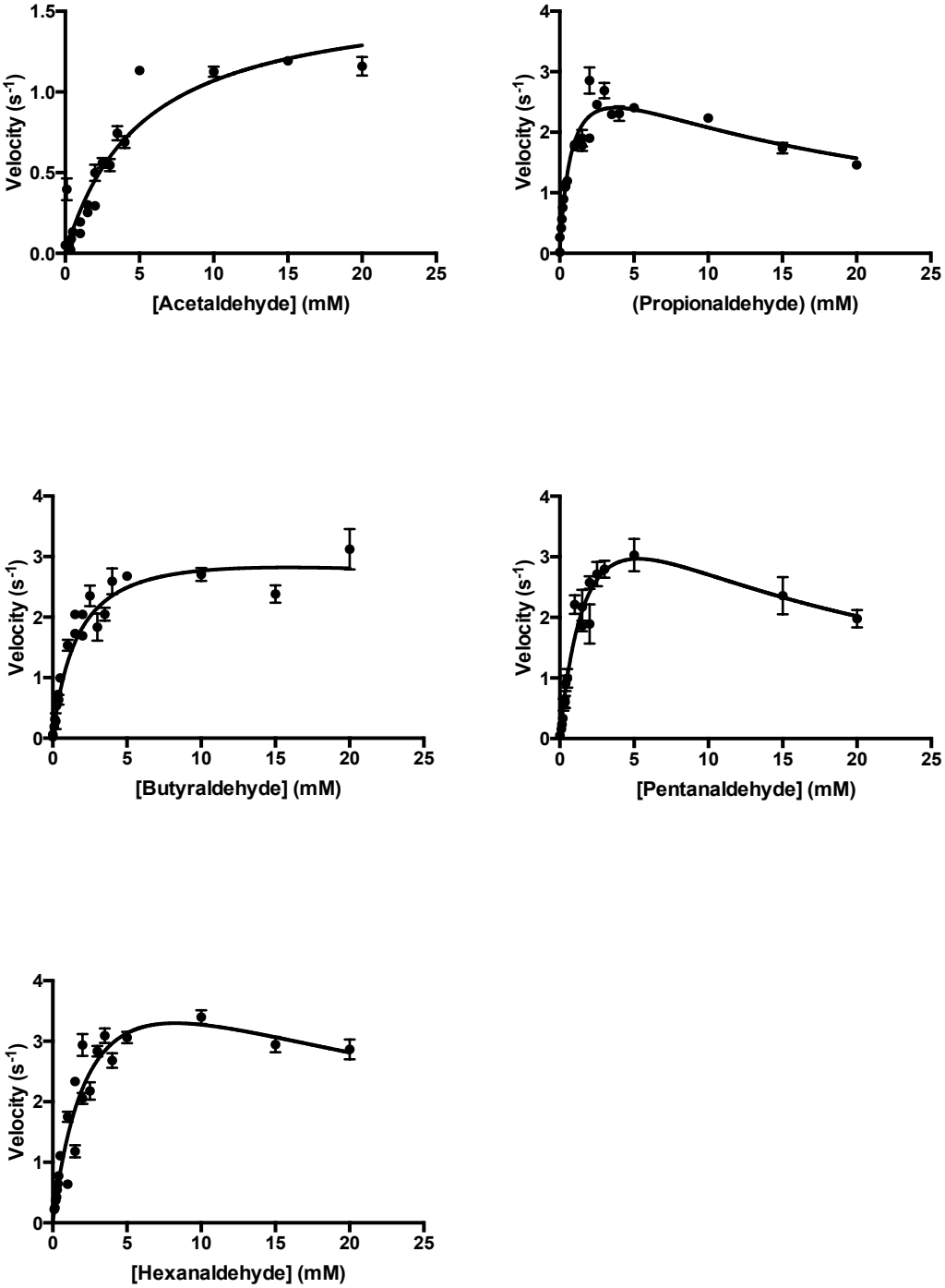


Figure 2.9 Activity of Cphy1178 against various aldehyde substrates.
Cphy1178 was incubated with increasing concentrations of aldehyde substrates in the following reaction mixture: 100 mM Tris.HCl (pH 8.0); 0.66 mM NAD^+, 100 mM KCl, 10 mM 2- mercaptoethanol. Data were analysed using Graphpad Prism and kinetic parameters were calculated by non-linear regression to the Michaelis-Menten equation with substrate inhibition for C3 and larger aldehydes. (Velocity is expressed as mM NADH s-1 mM-1ADH).

2.2.5 Expression and purification of the active site aldehyde dehydrogenase variant Cphy1178(20-462) (C269A and H387A)

Based on the identification of the potential catalytic cysteine (Cys269) and histidine (His387) residues from a sequence alignment of common aldehyde dehydrogenase homologues (described in detail in Chapter 3), these residues were mutated to an alanine residue, as this small, non-polar amino acid is unlikely to have any involvement in the acylation of an aldehyde substrate, which is the proposed mechanism of action for this enzyme. Mutants were generated using site-directed mutagenesis (described in Materials and Methods chapter), and were subject to both anion exchange and size-exclusion chromatography as in section 2.1.3. Figures 2.10 and 2.11 show the Size Exclusion Chromatography purification step for both variants. The Cphy1178 (C269A) and (H387) variants completely abolished NAD+ turnover activity, under the same experimental conditions as the wild-type protein. This suggests that these particular residues have an important role, at least in the initial NAD+ to NADH turnover step.

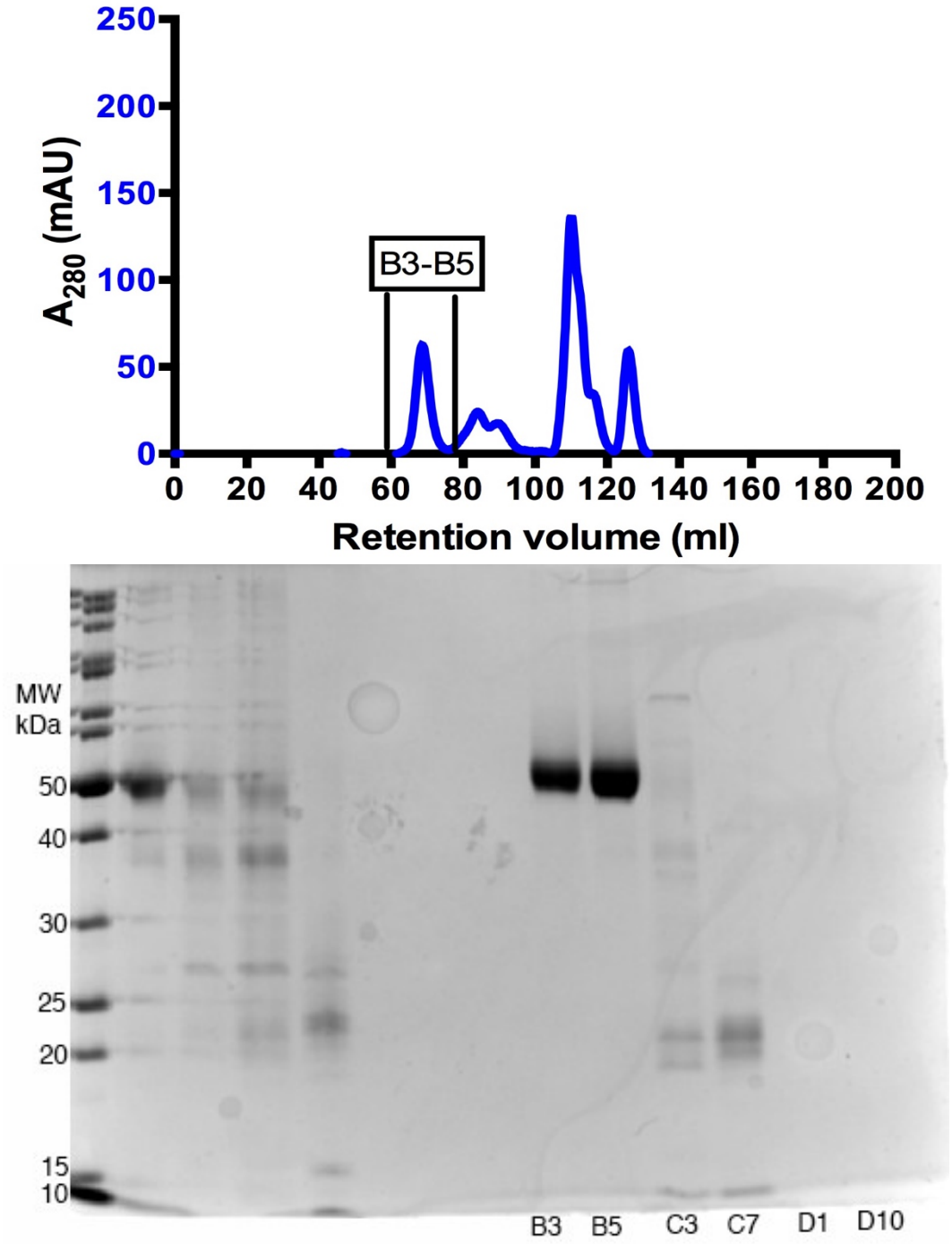


Figure 2.10 Size-exclusion chromatography purification of recombinant Cphy1178 (C269A) variant.
Fractions B3-D10 were analysed by 15% SDS-PAGE gel. Protein of the correct size was observed in fractions B3-B5

2.2.6 Summary and discussion

Our initial enzyme assays indicate that Cphy1178$_{20\text{-}426}$ is active against a range of short-chain fatty-aldehydes (C2-6). The other two proposed aldehyde dehydrogenases from the *Clostridium phytofermentans* genome were inactive under the conditions and against all the aldehydes tested. Two active site mutants, C269A and H387A were cloned via site-directed mutagenesis, recombinantly expressed and purified under the same conditions as the wild-type protein. The variants were also inactive against the tested short-chain fatty aldehydes, indicating that both residues have a role to play in the turnover of NAD+ to NADH in the cofactor-binding site. Further work to co-crystallise the wild-type Cphy1178 and both variants, should help further elucidate the mechanism of action of the enzyme

3 Crystal structure of an acylating aldehyde dehydrogenase from Clostridium phytofermentans, with Co-enzyme A and NAD+ in the active site

3.1. Introduction

In the C. phytofermentans genome, there are three putative aldehyde dehydrogenase enzymes associated with Bacterial Microcompartment loci: Cphy1178, Cphy1428 and Cphy2642 (Petit et al., 2013, 2015). A protein sequence alignment of these proteins and a homologue from the PDB (3K9D), as well as the putative aldehyde dehydrogenase from the *Clostridium difficile* ethanolamine utilisation locus (CD630_19170), *Clostridium beijerinckii* (CBS_3832) and *Salmonella enterica serovar Typhimurium* (STM2463) shows there are common conserved residues between the three proteins, most notably the catalytic cysteine (C269) and the active site histidine residue (H387) (Figure 3.1) Other sequence similarities are the glutamic acid (E357) residue, which is part of the NAD^+ cofactor-binding site, as well as some secondary structure elements, i.e. β-sheet 7 and α-helix 3 (Figure 3.1). Table 3.1 shows the sequence identity between the three putative aldehyde dehydrogenase enzymes from the *C. phytofermentans* BMCs, and the PDB structure from *Listeria monocytogenes*, as well as the other BMC aldehyde dehydrogenases mentioned above. Interestingly, there is a higher amino acid sequence identity between Cphy1178 and aldehyde dehydrogenases from other species (highest Cphy1178 and 5JFN – 54%) than between Cphy1178 and the other putative AldDHs from *C. phytofermentans*, Cphy1428 and Cphy2642 (30% and 31% respectively). The highest level of sequence identity is between 3K9D from *L. monocytogenes* and Cphy2642. The higher level of sequence identity between Cphy1178 and the 5JFN PDB homologue could be

explained by the fact that this enzyme is also a propionaldehyde dehydrogenase, which Cphy1178 was shown experimentally to be in Chapter 2.

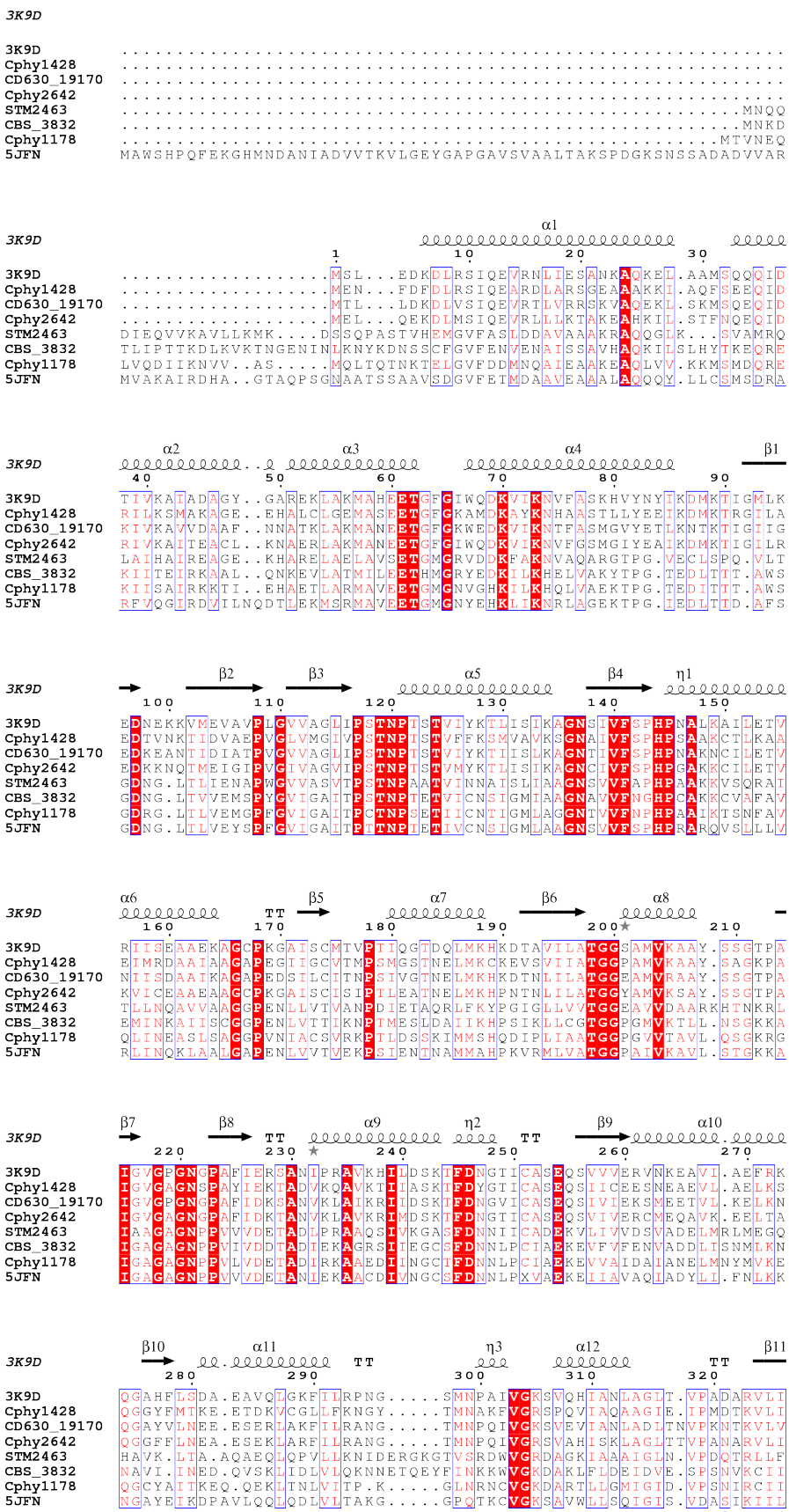
	Cphy1178	Cphy1428	Cphy2642	3K9D	CBS_3832	STM2463	CD630_19170	5JFN
Cphy1178	100							
Cphy1428	30	100						
Cphy2642	31	49	100					
3K9D	30	49	66	100				
CBS_3832	52	28	30	28	100			
STM2463	42	31	33	31	44	100		
CD630_19170	30	50	64	64	28	32	100	
5JFN	54	31	31	32	51	45	32	100

Table 3.1 Amino acid sequence identity (%) between aldehyde dehydrogenases from different bacterial species.
Sequence identity was calculated using a BLASTp alignment. Protein sequences were taken from KEGG or from the protein PDB entry

The aldehyde dehydrogenase enzymes form a key part of the sugar breakdown pathways in BMCs in various bacterial species (Yeates et al., 2013), as they turn over the reactive aldehyde intermediates within these pathways to generate useful acyl-CoA derivatives (Leal et al., 2003). The AldDH enzymes have been shown to have a Rossmann-fold nucleotide binding domain that positions an $NAD(P)^+$ in the suitable conformation for hydride transfer from the aldehyde substrate (Liu et al., 1997)

Although there is a glutamic acid residue (E357) in the NAD^+ active site, it is not in the correct conformation to act as a general base to resolve the acyl-enzyme intermediate to the free acid, and so it is likely that another cofactor, Co-enzyme A (CoA) is used to resolve this intermediate (Shone and Fromm, 1981). Up until now there were no structures of an acylating aldehyde dehydrogenase with the CoA cofactor shown in the active site. This chapter looks at the X-ray crystallographic

structures of Cphy1178$_{(20-462)}$, the C269A and H387A variants with NAD^+ and CoA in the cofactor binding site, and a final active site variant, E357A, to confirm the overall aldehyde dehydrogenase mechanism


3K9D
Cphy1428
CD630_19170
Cphy2642
STM2463
CBS_3832
Cphy1178
5JFN

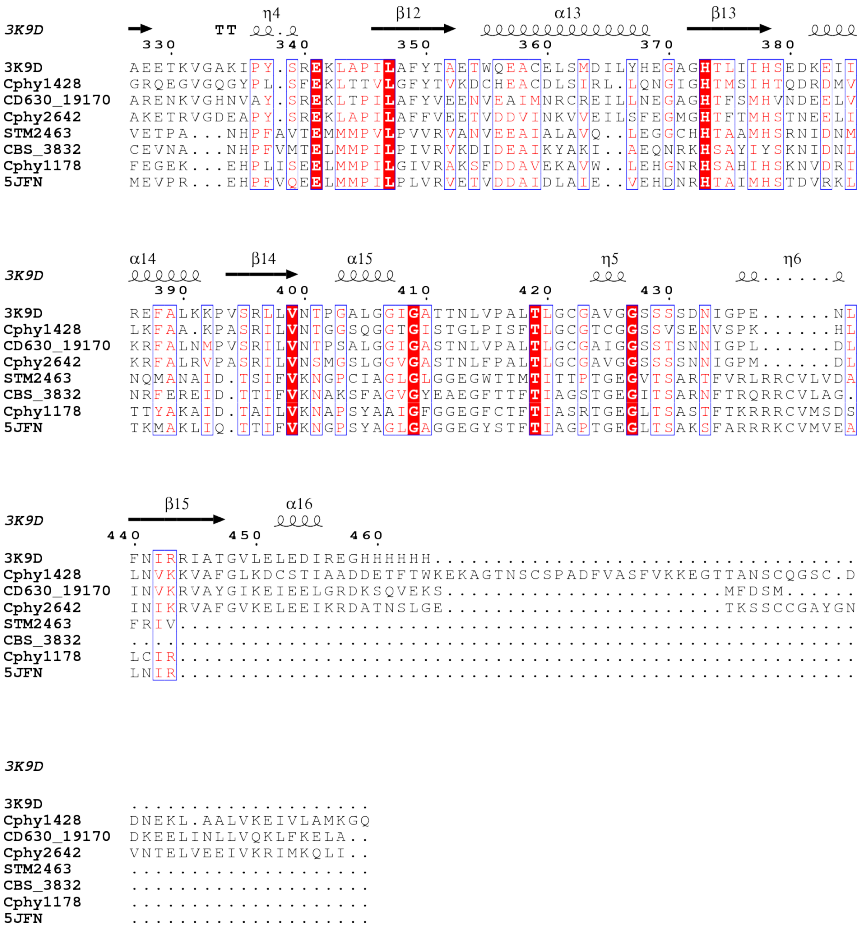


Figure 3.1 Multiple sequence alignment of aldehyde dehydrogenase enzymes used in this study and close homologues.
Sequence alignment generated using ClustalΩ (Sievers and Higgins, 2014) and displayed using ESPript(Gouet et al., 2003). Secondary structure elements of Cphy_1178 are shown above the alignment.

3.1.1 Crystallisation of Cphy1178(20-462) and active site variants

For crystallisation trials, the untagged and truncated variant for each construct was purified by anion exchange and the size exclusion chromatography (described in Chapter 2) and was concentrated to between 8-12 mg/ml, with protein concentration estimated based on calculated extinction coefficients at 280 nm. Proteins were initially screened for crystallisation against a range of commercial screens and optimised around conditions from the JCSG+ suite (Materials and Methods). Once suitable crystallisation conditions had been found for both Cphy1178$_{(20\text{-}462)}$ and the active site

variants (0.1 M sodium acetate pH4.7, 1.8 M ammonium sulfate), crystals were harvested, and data collected and refined as described in Materials and Methods. Table 3.2 shows the X-ray data and refinement statistics. Figure 3.2 shows the bi-pyramidal crystals grown from the optimisation trays.

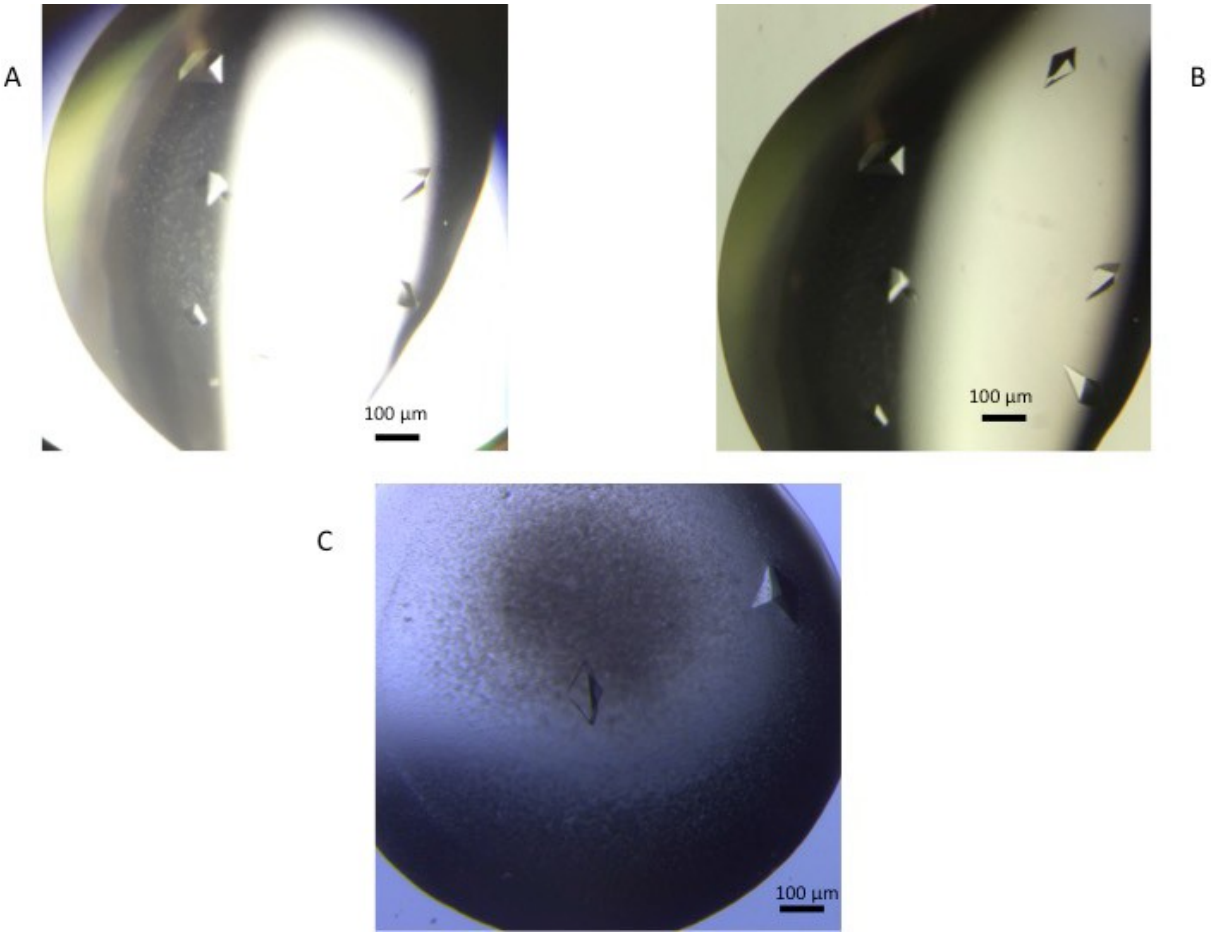


Figure 3.2 Images of bi-pyramidal Cphy1178$_{(20-462)}$ and variant crystals.
Images taken with a Leica DFC450 C digital microscope at 10X resolution from Linbro 24-well optimisation plates A)Cphy1178$_{(20-462)}$, crystals B) and C) are the C269A variant

	Cphy1178(20–462)			
	NAD+	**C269A-CoA**	**H387A**	**E357A**
Data Collection				
Wavelength (Å)	0.9795	0.9795	0.9700	
Resolution range (Å)	32.64–1.64 (1.67–1.64)	49.89–1.77 (1.8–1.77)	40.53–2.083 (2.158–2.083)	49.8 - 1.851 (1.917 - 1.851)
Space group	I 41 2 2	I 41 2 2	I 41 2 2	I 41 2 2
Unit cell (Å)	138.50, 138.50, 84.60	138.30, 138.30, 84.45	138.45, 138.45, 84.75	137.86, 137.86, 84.45
Total reflections	299,010 (14,770)	518,921 (25,397)	24,2393 (23,123)	50,7886 (47,875)
Unique reflections	50,131 (24,674)	40,079 (2229)	24,570 (2,389)	34835 (3415)
Multiplicity	6.0 (6.0)	12.9 (11.4)	10.0 (9.7)	14.6 (14.0)
Completeness (%)	99.7 (99.9)	99.9 (98.5)	98.85 (99.25)	99.95 (99.94)
Mean I/sigma(I)	17.8 (2.3)	22.4 (2.3)	18.94 (1.63)	24.08 (1.41)
Wilson B ($Å^2$)	16.52	24.95	40.85	33.57
Rmerge	0.062 (0.697)	0.073 (1.050)	0.081 (1.227)	0.08105 (1.889)
Rmeas	0.074	0.079 (1.152)	0.08496	0.084 (1.96)
CC1/2	0.999 (0.744)	0.999 (0.738)	1 (0.649)	1 (0.644)
Diffraction images (DOI)	10.7488/ds/1307	10.7488/ds/1318	10.7488/ds/1319	Not published
Model Building and Refinement				
Rwork	0.1463 (0.2110)	0.1465 (0.2122)	0.1823 (0.2651)	0.1719 (0.2598)
Rfree	0.1738 (0.2376)	0.1830 (0.2682)	0.2575 (0.3236)	0.2161 (0.3041)
Number of non-hydrogen atoms	3795	3597	3376	3659
macromolecules	3293	3225	3245	3277
ligands	69	79	5	78
water	433	293	126	242
Protein residues	435	431	435	439
RMS(bonds) (Å)	0.011	0.010	0.015	0.011
RMS(angles) (°)	1.37	1.26	1.44	1.34
Ramachandran				
favored (%)	99	99	98	98.38
outliers (%)	0	0	0	0
Clashscore	2.66	0.45	4.42	3.56
Average B-factor ($Å^2$)	24.90	33.90	56.80	38.60
macromolecules	23.90	32.50	57.10	37.27
ligands	28.30	79.8	80.8	85.51
solvent	32.10	37	46.50	41.54
PDBID	4C3S	5DBV	5DRU	Not deposited

Table 3.2 X-ray data collection and refinement statistics for Cphy1178 protein and active-site variants.
Statistics for the highest-resolution shell are shown in parentheses.

3.1.2 Structure of Cphy1178$_{(20-462)}$

To understand the co-factor binding properties of Cphy1178, the structure of an N-terminal truncation was determined, comprising residues 20 – 462, with bound NAD^+ cofactor to 1.6 Å resolution (Tuck et al., 2016). In these structures, the asymmetric unit contains a single chain with residues 28 to 462 visible in the electron density map. The overall structure corresponds to other members in the aldehyde dehydrogenase family (Figure 3.3A), with a catalytic domain (residues 238 – 427, highlighted pink), a co-factor-binding domain with a Rossmann-fold nucleotide binding architecture (residues 28 – 108, 127 – 237, and 428 – 447, highlighted grey), and an oligomerisation domain (residues 109 – 126 and 448 – 462, highlighted green). The catalytic and nucleotide binding domains come together to form an extended nucleotide and ligand-binding tunnel that is open at both ends, with the catalytic cysteine (C269) at the centre of the tunnel and the NAD^+ cofactor binding at one side (Figure 3.3B). The ligand-binding tunnel is about 5 Å in diameter at its widest point and spans 16 Å from the solvent exposed entry point to the catalytic cysteine and is lined with hydrophobic residues. The tunnel is long enough to accommodate up to a C10 aldehyde, although in vivo the enzyme is unlikely to encounter such a substrate.

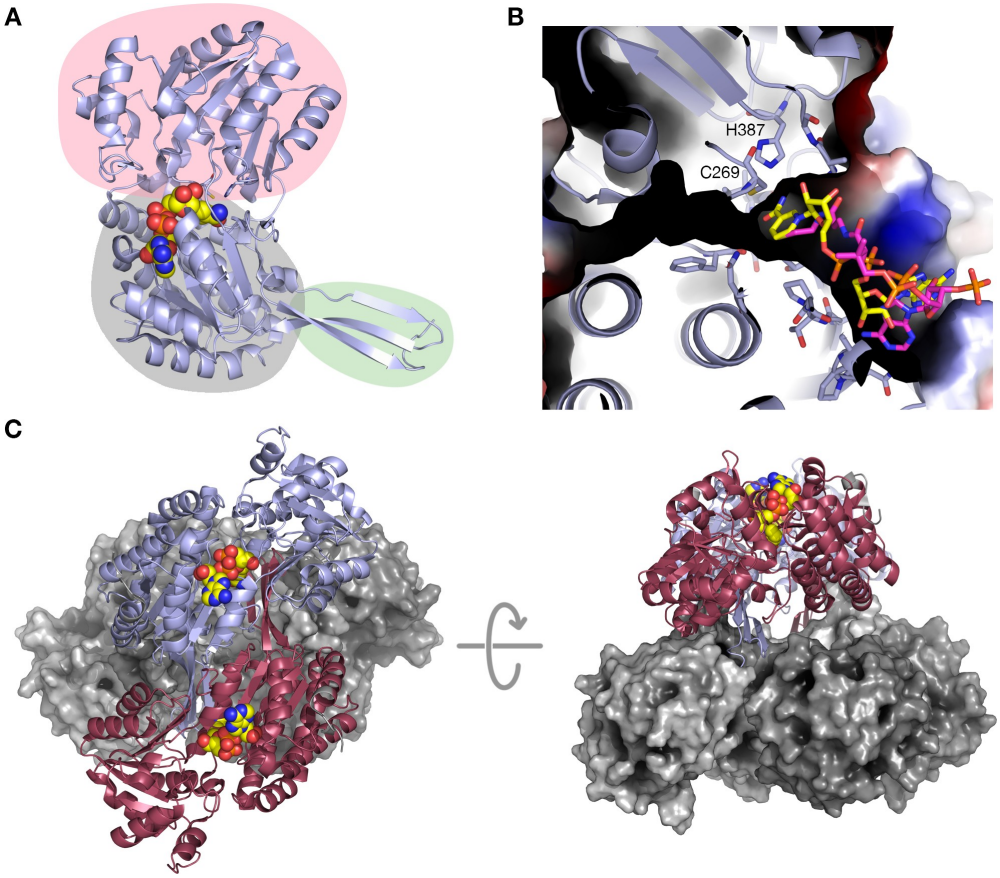


Figure 3.3 Structure of Cphy1178$_{(20-462)}$
A) Secondary structure cartoon of a monomer of Cphy1178$_{(20-462)}$ showing bound NAD^+ as spheres coloured yellow for carbon, red for oxygen, blue for nitrogen and orange for phosphorus. The catalytic domain is highlighted pink, Rossmann-fold nucleotide binding domain in grey and oligomerisation domain in green. B) Active site tunnel and nucleotide binding pocket. The surface of Cphy1178$_{(20-462)}$ is shown coloured by electrostatic potential (blue for positive, red for negative) and the secondary structure cartoon is shown in blue. The catalytic cysteine and histidine residues are shown as sticks. Bound cofactors, NAD^+ and CoA are shown with yellow and pink carbons respectively. C) Cphy1178 forms a dimer of dimers quaternary structure. Two subunits are shown as cartoons, with NAD^+ shown as spheres to show the relative orientation of the nucleotide-binding cleft; two further subunits are shown as grey surface representations (Tuck et al., 2016)

3.1.3 Structure of Cphy1178(20-462) C269A and H387A variants

The structure of Cphy1178 was also determined in a number of active site variant forms: C269A with bound CoA (1.77 Å resolution) and H387A (2.08 Å resolution) to ensure that mutation of these residues did not destabilise the structure of the protein and to probe the active site. Both proteins displayed essentially identical structures to the wild-type protein. In the structure of the Cphy1178C269A variant the position of

the loop between 335 and 339 was not clear in the electron density maps, so was omitted from the final structure refinement.

3.1.4 Co-factor binding

The structure of Cphy1178$_{(20-462)}$ was determined with NAD^+ bound in the nucleotide binding cleft and active-site pocket after soaking apo-crystals with crystallization solution supplemented with 10 mM NAD^+. The ligand displayed excellent electron density (Figure 3.5A) and was refined to an occupancy of 0.83 in the final structure using phenix.refine. The NAD_+ is found in the hydride transfer conformation (Bateman et al., 2003; Perez-Miller and Hurley, 2003) and participates in 4 direct hydrogen bonding interactions with the enzyme and a number of other interactions mediated by sulfate ligands.

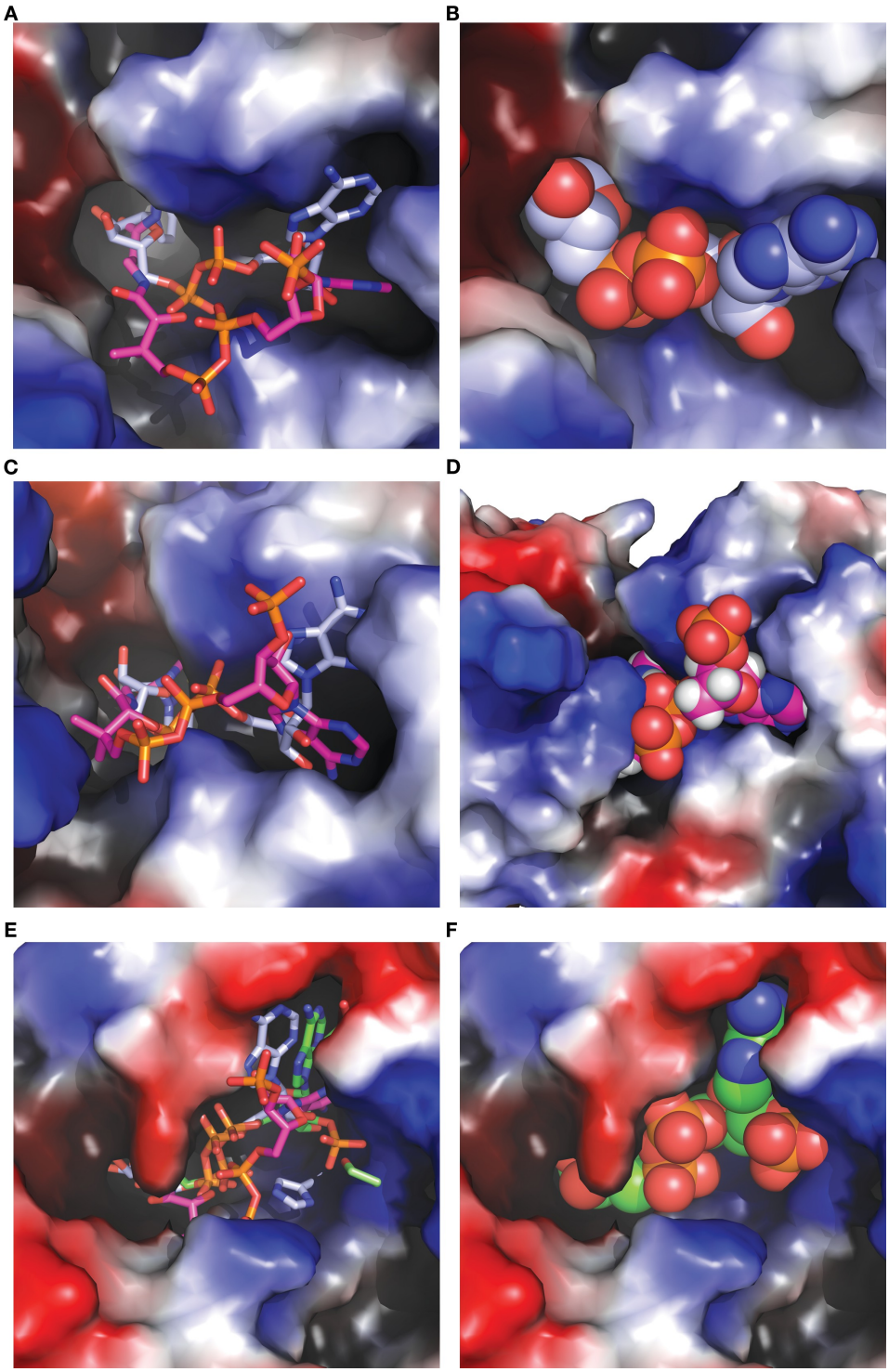


Figure 3.4 Surface views of cofactor binding to Cphy1178$_{(20\text{-}462)}$.
A) Comparison of co-factor positions in Cphy1178(20-462) with $NADP^+$ position in PDBID:4NMJ. Electrostatic surface shown for 4NMJ and carbon atoms of $NADP^+$ shown in blue. B) Space- filling view of $NADP^+$ bound to 4NMJ. C) Top view of adenine binding pockets in Cphy1178$_{(20\text{-}462)}$. NAD^+ shown with blue carbon atoms, CoA shown with magenta carbons. D) Space-filling representation of NAD^+. E) Oblique view of adenine binding pockets depicted as in C). F) Space-filling representation of CoA (Tuck et al., 2016).

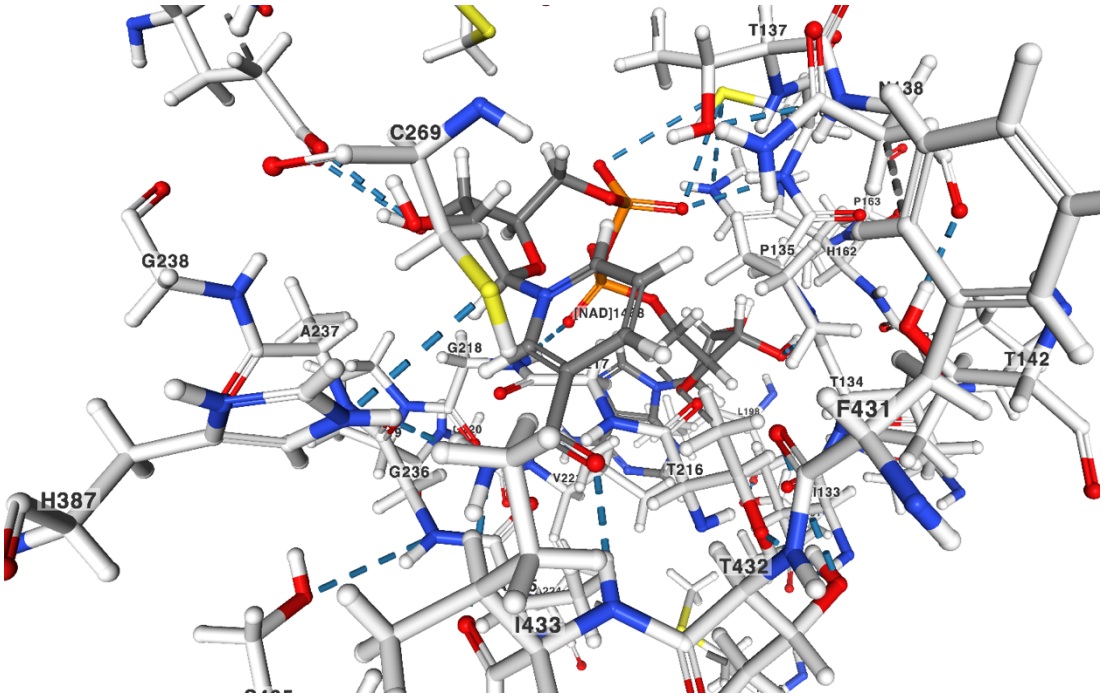


Figure 3.5 Stick representation of NAD+ binding in the cofactor binding site.
Hydrogen bonding interactions shown as a blue dashed line. NAD+ molecule in grey and Cphy1178(20-264) shown in white. Blue circles highlight the catalytic C269 residue, the nicotinamide ring and the I433 residue. PDB ID 4C3S and (Tuck et al., 2016), NGL Viewer (Rose et al., 2018) NGL viewer: web-based molecular graphics for large complexes. Bioinformatics and RCSB PDB

The adenine ring does not directly participate in any hydrogen bonding interactions, but it is positioned between Leucine 198 and Valine 221. The N7 of the adenine ring accepts a hydrogen bond from an ordered solvent molecule that is also coordinated by an oxygen atom in the adenosine phosphate group and the backbone nitrogen of Valine 221. Both O2 and O3 of the adenosine ribose interact with an ordered water molecule that bridges them to the carbonyl oxygen of Proline 161. Similarly, two oxygen atoms from the phosphate groups are bridged to Histidine 162 via a solvent molecule. Glutamine 357 participates in hydrogen bonding interactions with O2 and O3 of the nicotinamide ribose. The nicotinamide ring is positioned next to the catalytic Cysteine

269 through hydrogen bonding interactions between N7 and O7 and the peptide backbone of Isoleucine 433 (Figure 3.5). The absence of enzymatic activity in the presence of the $NADP^+$ co-factor is explained by the presence of a histidine and proline residue blocking the position that the 2'-phosphate adopts in structures of aldehyde dehydrogenase enzymes determined with NADP+ bound (Figure 3.6A/B).

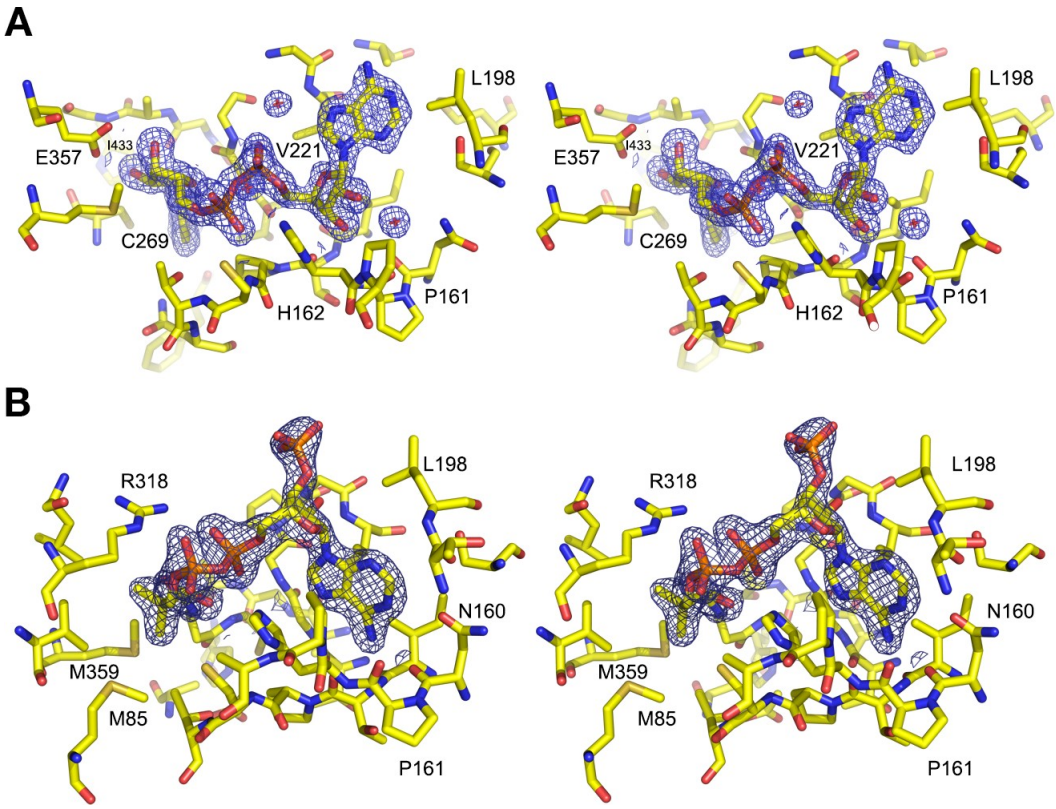


Figure 3.6. Wall-eyed stereo view of electron density maps of bound NAD+ and CoA cofactors. A) Structure of NAD^+ soaked $Cphy1178_{(20-462)}$. NAD^+ and protein residues within 4 Å shown as stick representations and 2mFo-DFc map shown as a blue mesh contoured at 1σ. B) Structure of CoA soaked $Cphy1178_{(20-462)}$C269A displayed as in A) (Tuck et al., 2016).

The structure of $Cphy1178_{(20-462)}$ C269A was determined using crystals soaked with 10 mM CoA. The cofactor showed good electron density for the adenosine group, but poor density for the pantothenic acid group, particularly the terminal region with the sulfydryl group in the active site, which has no visible electron density (Figure 3.6B). A comparison of the structure of the NAD^+ and CoA bound forms of the protein show

that the structures are essentially identical, with an rmsd Cα of 0.24 Å over 431 residues. The two structures differ only in the position of the loop between residues 215 and 223, where P219, G220 and V221 are shifted toward the adenine ring in the NAD^+ -bound structure (Figure 3.7). In contrast to the models of CoA binding to the *Pseudomonas* DmpF aldehyde dehydrogenase (Lei et al., 2008) and *Lactobacillus reuteri* PduP aldehyde dehydrogenase (Sabet-Azad et al., 2013) this crystal structure shows distinct adenine-binding modes for NAD^+ and CoA . The adenine ring is flipped by 180 degrees relative to its position in the NAD^+ structure and makes hydrogen-bonding contacts with both the side chain and carbonyl oxygen of Asparagine 160, and the carbonyl oxygen of Threonine 134, placing the phospho-ribose group above both of the nucleotide binding pockets. The phosphate groups between the nucleotide and pantothenate group are held in place by hydrogen bonds formed between the side-chain nitrogen atoms of Histidine 162 and Arginine 318. While there is clear electron density for the nucleotide and phosphate-proximal portion of the pantothenate group of the CoA the β-mercapto- ethylamine group is not visible, indicating that the distal arm of CoA is flexible in the absence of thio-acyl substrates bound to the catalytic cysteine for acyl transfer.

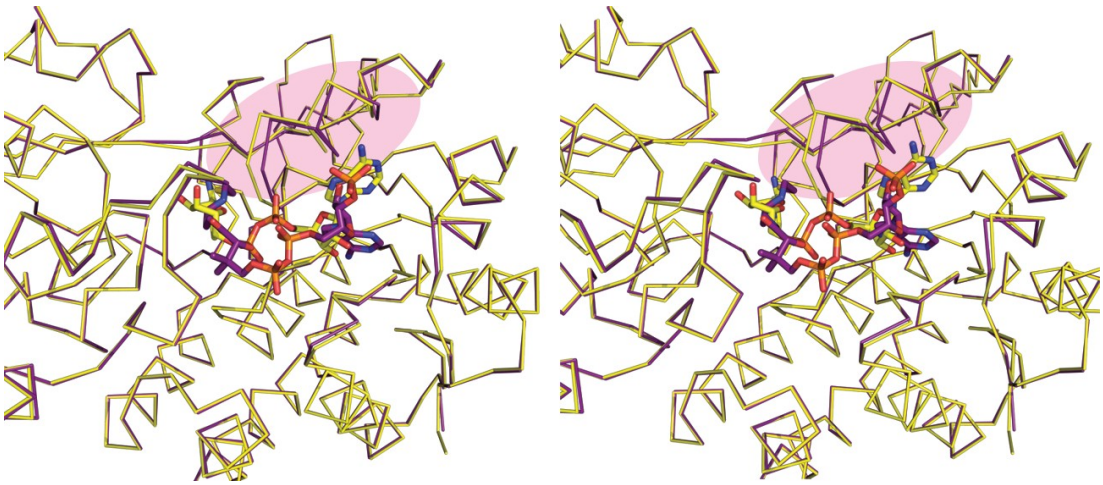


Figure 3.7 Wall-eyed stereo view of cofactor binding within the nucleotide- binding domain of Cphy1178$_{20-462}$.
Stereo ribbon representation of the proteins and stick representations of the NAD+ and CoA cofactors are displayed to show the minor difference between the two cofactor bound structures of Cphy1178(20-462). The helix highlighted in pink, between residues 215 and 225 is shifted slightly in the CoA bound structure (purple) when compared to the NAD+ bound structure (Tuck et al., 2016).

3.1.5 Proposed mechanism of action of Cphy1178(20-462)

We have shown that Cphy1178 shows a higher level of activity with propionaldehyde as a substrate compared to other aldehydes tested and is capable of acyl-transfer to CoA using this substrate. We have also determined the structure of this enzyme with CoA bound in the active site. The acylating aldehyde dehydrogenase family enzymes do not possess the glutamic acid general-base that is required to activate the catalytic cysteine and to deprotonate water for deacylation of the acyl-enzyme intermediate. In Cphy1178 this residue is replaced by an alanine (A235) and is a small hydrophobic residue in all of the acylating aldehyde dehydrogenases shown in Figure 3.1. Mutagenesis of a strictly conserved Histidine residue (H387) (Figure 3.1), completely abolishes activity of the enzyme. This residue is not found in any of the non-acylating aldehyde dehydrogenase enzymes. The proximity of this residue to the catalytic cysteine is incompatible with the formation of the correct geometry for deprotonation

of water for resolution of the acyl-intermediate to produce a carboxylic acid product. The biochemical data showing that the H387A variant is inactive, suggests that this residue acts as a base to activate the catalytic cysteine and to stabilise the acyl-transfer intermediate between the enzyme and CoA cofactor. Due to the proximity of a glutamic acid residue (E357) to the cysteine group of CoA and the strict conservation of this residue, a model is proposed in which this residue activates the CoA cofactor for acyl transfer from the acyl- enzyme intermediate (Figure 3.8). The reaction scheme shows H387 deprotonates the catalytic cysteine to allow nucleophilic attack on the substrate to form tetrahedral intermediate. Consistent with previous reports the oxyanion would be N138. After hydride transfer to NAD^+, the NADH product leaves the cofactor-binding pocket in the rate-limiting step, followed by entry of CoA, which is deprotonated by E357 and subsequently attacks the acyl-enzyme intermediate to produce the thioester product and free enzyme (Figure 3.8). The extra proton on E357 is likely transferred to bulk water via a proton-relay system comprising the conserved residues N138 and K94.

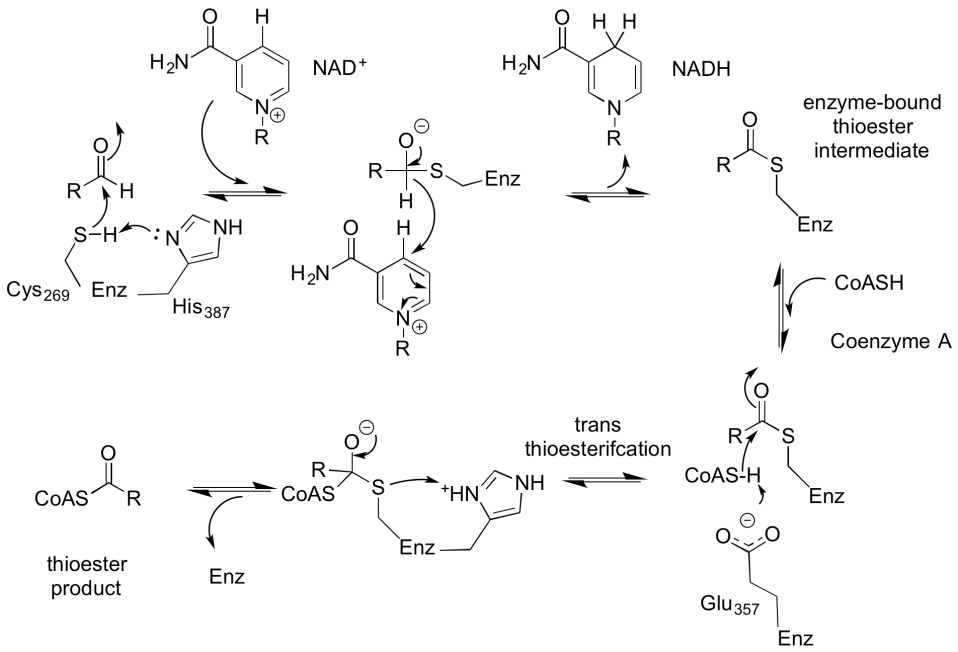


Figure 3.8 Proposed catalytic mechanism of Cphy1178.
The catalytic cycle proceeds via a bi-uni-uni-uni-ping-pong mechanism involving acylation of the catalytic cysteine (C269) and hydride transfer to NAD+ , followed by trans-thioesterification between the enzyme and CoA to produce the acyl-CoA product. A ping-pong mechanism, also known as a double displacement reaction, is characterised by the enzyme changing to an intermediate form during the first substrate to product reaction (Tuck et al., 2016). The bi- and uni- prefixes refer to the number of substrates and products, so two and one respectively (Cleland, 1963).

Interestingly, a recent paper from (Zarzycki et al., 2017) shows the crystal structure of an acylating prionaldehyde dehydrogenase from *Rhodopseudomonas palustris*, with CoA or propionyl-CoA in the active site (Zarzycki et al., 2017). Unlike in the mechanism proposed here, they observed that the S-propionylcysteine assumes a different rotamer confirmation than the non-acylated cysteine, which prevents the interaction of the CoA with the conserved glutamate residue (E357). They suggest that two conserved water molecules are responsible for the deprotonation of CoA. To test whether the glutamic acid residue E357 is responsible for the deprotonation of CoA, a variant of the Cphy1178(20-462) was generated, replacing the glutamate residue with an alanine residue. The variant was expressed, purified and crystallised as in described in Chapter 2 and earlier in this chapter.

3.1.6 Structure of Cphy1178(20-462) E357A variant

The structure of the E357A variant was almost identical to the wild-type version of the protein. Interestingly, although the crystals were not soaked with the CoA cofactor, there was electron density in the structure that fitted for the adenine ring (Figure 3.9 and 3.10). This suggests that CoA has been trapped within the active site, as it has not been able to deprotonate due to the lack of glutamic acid residue to carry out the acyl-transfer step. The conserved water molecules described in Zarzycki et al. (2017), were also not observed, leading back to the proposed mechanism of the conserved glutamic acid residue deprotonating the CoA thiol.

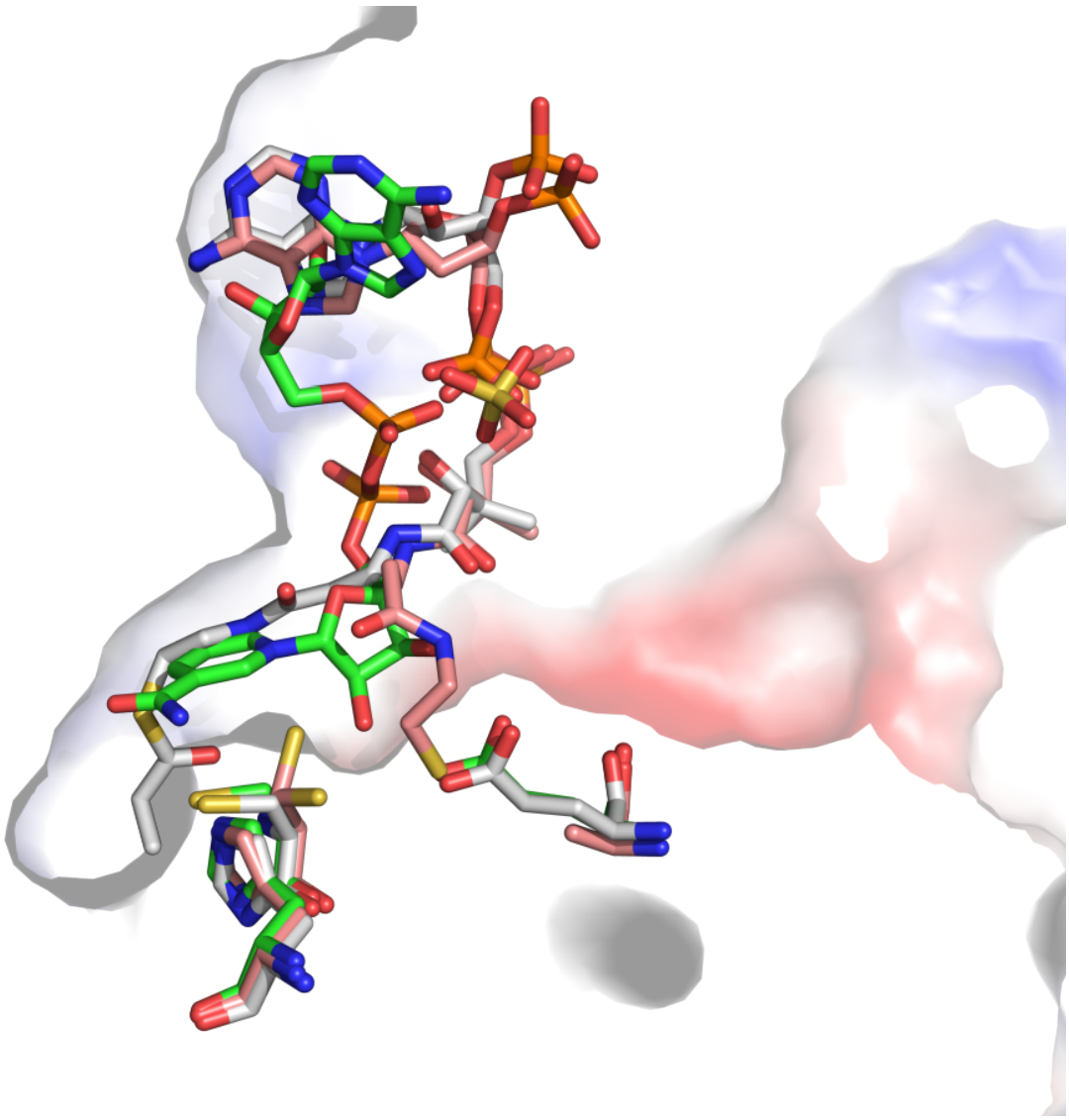


Figure 3.9 Architecture of cofactor binding to Cphy1178(20-462).
Surface slice through Cphy117820–462 showing the position of NAD+ from (PDB ID 4C3S) (green), the CoA from *Rhodopseudomonas* palustris (PDB ID 5JFN) and the CoA from Cphy1178 (E357A) (pink). Protein residues shown as sticks in the co-factor binding tunnel, shown as an electrostatic surface.

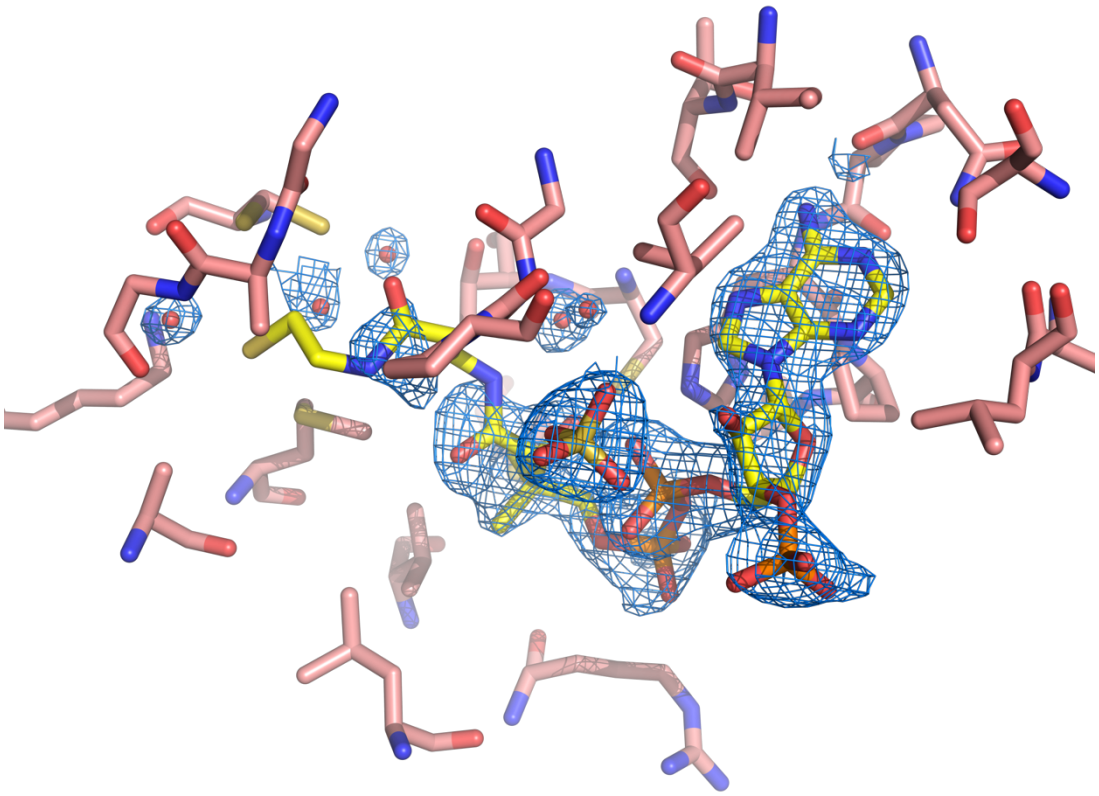


Figure 3.10. Electron density map of potential bound CoA cofactor
Structure of CoA (yellow sticks) modelled in the Cphy1178(20-462).E357A variant (pink sticks). Enzyme and CoA shown as stick representations and 2mFo-DFc map shown as a blue mesh contoured at 1σ.

3.1.7 Summary and discussion

The crystal structures of Cphy1178(20-462) (4C3S) and the variants C269A (5DBV), H387A (5DRU) and E357A (undeposited) were solved with either NAD+ or the Co-enzyme A co-factor modelled in the Rossmann-fold nucleotide binding domain. The active site variants were almost identical to the wild-type protein structurally; however the C269A and H387A variants were found to be inactive (Chapter 2). This data helped to propose a mechanism of action for the enzyme, whereby the H387 deprotonates the catalytic cysteine to allow nucleophilic attack on the substrate to form a tetrahedral intermediate. After hydride transfer to NAD^+, the NADH product leaves the cofactor-binding pocket in the rate-limiting step, followed by entry of CoA, which is

deprotonated by E357 and subsequently attacks the acyl- enzyme intermediate to produce the thioester product and free enzyme. Along with the biochemical data shown from Chapter 2, Cphy1178 is proposed to be a propionaldehyde dehydrogenase and confirms the reaction scheme proposed in (Petit et al., 2013)(Figure 3.11).

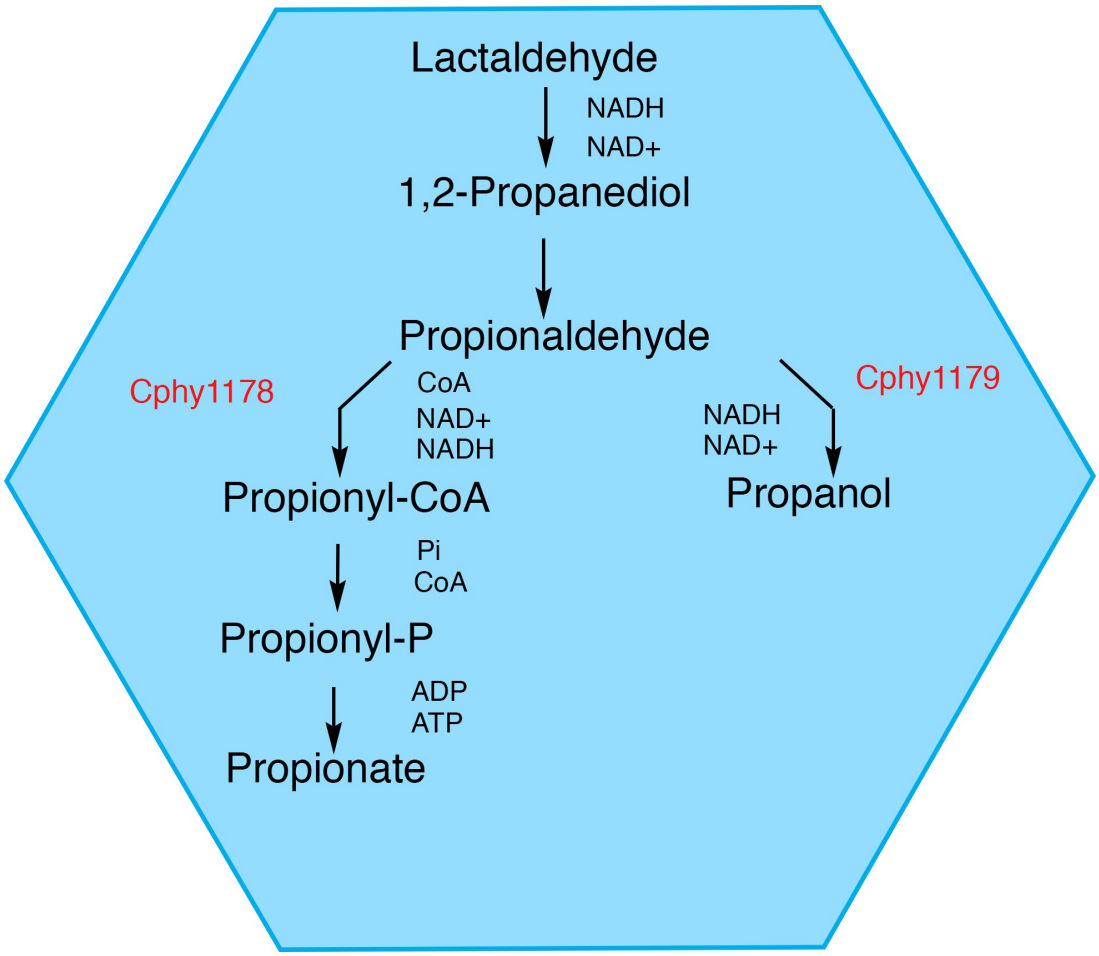


Figure 3.11 Scheme for Clostridium phytofermentans fucose/rhamnose utilisation pathway. Pathway intermediates and cofactors shown with proposed aldehyde dehydrogenase (Cphy1178) and alcohol dehydrogenase (Cphy1179) shown in red. Scheme based on (Petit et al., 2013).

4 Bacterial microcompartments as tools for synthetic biology

4.1 Introduction

Although there are conflicting ideas on the official definition of synthetic biology, a general view is that it is 'the rational design of biological systems and living organisms using engineering principals' (Osbourn et al., 2012). Bacterial microcompartments have certain properties that make them potentially useful tools in synthetic biology, raising solutions to common bioengineering problems.

They are modular, have compact organization, and are found in a wide variety of ecological niches including plant-associated and pathogenic bacterial species, so could be relevant in both the development of biofuels (Machado and Atsumi, 2012), as well as a target in the battle against antibiotic resistance (Chowdhury et al., 2014).

With the emergence of tools for synthetic biology, it is possible to generate large DNA libraries quickly and cost-effectively. Standard cloning can be time-consuming and is a main stumbling block in a molecular laboratory workflow; however, the last decade has seen the development of simple one-pot restriction/ligation methods based on Type IIs restriction endonucleases that cut outside of their recognition sites, such as the Golden Gate method (Engler et al., 2008)A number of standards for DNA assembly and part design have been produced in the last few years (Engler et al., 2009, 2014; Moore et al., 2016; Sarrion-Perdigones et al., 2011; Weber et al., 2011; Werner et al., 2012) including the CIDAR Modular Cloning toolkit (MoClo), developed by the Densmore laboratory in Boston (Iverson et al., 2015). The toolkit, which was the first bacterial MoClo parts library is available publicly through Addgene. This kit consists of a series of plasmids containing different strength promoters, ribosome binding sites

(RBSs), open reading frames (ORFs) for a range of fluorescent reporter genes, terminators and plasmids for transcriptional units. The premise of this toolkit is described in detail in Chapter 1.5; however, it is important to note that the flexibility offered by this kit is central to this study into building bacterial microcompartments as tools for synthetic biology. The iterative synthetic biology workflow, illustrated in Figure 4.1: of designing, building, testing, learning and using the constructs generated, were used in the following parts of this project.

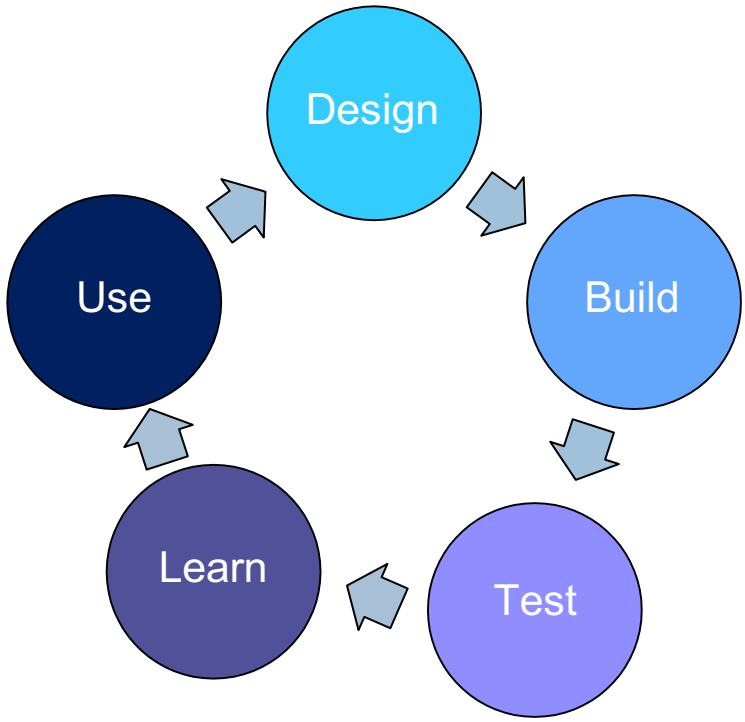


Figure 4.1. The synthetic biology workflow (Myers et al., 2017).

The aims are:

1. To use the CIDAR MoClo toolkit to construct a shell protein library to develop various transcriptional units with different numbers of shell proteins from different bacterial species, with a view to expressing and visualizing these synthetic microcompartment shells with thin-section electron microscopy and ultimately using them as fully functional synthetic BMCs .

2. To design forward engineered ribosome binding site and localisation sequence fusion-constructs using the BMC localisation sequences from the Salmonella and C. phytofermentans Pdu microcompartments. The rationale behind this is to find out if a fixed protein part downstream of the different strength RBSs will lead to differing expression levels of fluorescent reporter, or if there is a particular strength RBS that will be optimal for production of the reporter. This can be used in further studies to test the targeting of a specific amount of protein to a microcompartment.

3. To use the Golden-Gate modified pET28a vector to clone and express the EutM shell proteins from different bacterial species. Overexpressed shell proteins can either be crystallised to look at the surface residues and charge more closely, or the structures can be shown by thin-section microscopy

4.2 Shell protein library construction using CIDAR MoClo toolkit

4.2.1 Introduction

The different sizes and charges of the bacterial microcompartment shell proteins can greatly affect microcompartment formation, as can the differing charge states around the pore allowing for influx of different BMC pathway molecules (Crowley et al., 2008; Pitts et al., 2012; Zarzycki et al., 2015). Efforts have been made to rationally design synthetic microcompartment shells for various uses, although the general idea is to target novel enzyme pathways to microcompartment shells, for either increased enzyme efficiency or to sequester toxic, or volatile, molecules away from the bacterial cell (Cai et al., 2015, 2016; Lassila et al., 2014; Parsons et al., 2010; Slininger Lee et al., 2017).

4.2.2 Shell proteins in Clostridium phytofermentans fucose/rhamnose and Salmonella Pdu BMCs

There are six putative bacterial microcompartment shell proteins encoded on the *Clostridium phytofermentans* fucose/rhamnose BMC locus, and seven on the *Salmonella* Pdu BMC locus. The Pfam domains of each of these proteins are shown in Table 4.1, along with the ethanolamine utilisation microcompartment homologues.

Shell protein	Pfam domain	Protein homologue(s)
Cphy_1176	00936	EutS
Cphy_1180	00936	-
Cphy_1181	00936	-
Cphy_1182	00936	-
Cphy_1184	03319	EutN/CcmL
Cphy_1186	00936-00936	-
STM2038 (PduA)	00936	EutM
STM2039 (PduB)	00936-00936	-
STM2045 (PduJ)	00936	EutM
STM2046 (PduK)	00936	-
STM2049 (PduN)	03319	EutN/CcmL
STM2054 (PduT)	00936-00936	-
STM2055 (PduU)	00936	EutS

Table 4.1 BMC shell proteins from *Clostridium phytofermentans* and *Salmonella enterica* used in this study.
The protein homologues were found on the KEGG GENE database.

The BMC shell proteins fall into three distinct domain families, BMC-H, BMC-T and BMC-P (Lassila et al., 2014). The majority of the shell proteins are BMC-H type domains, indeed this is the most common type among all BMCs so far. A sequence alignment of the BMC-H–type proteins (Pfam 00936) against the EutM homologue (PDBID:4AXJ) from *Clostridium difficile* (Pitts et al., 2012)is shown in Figure 4.2.

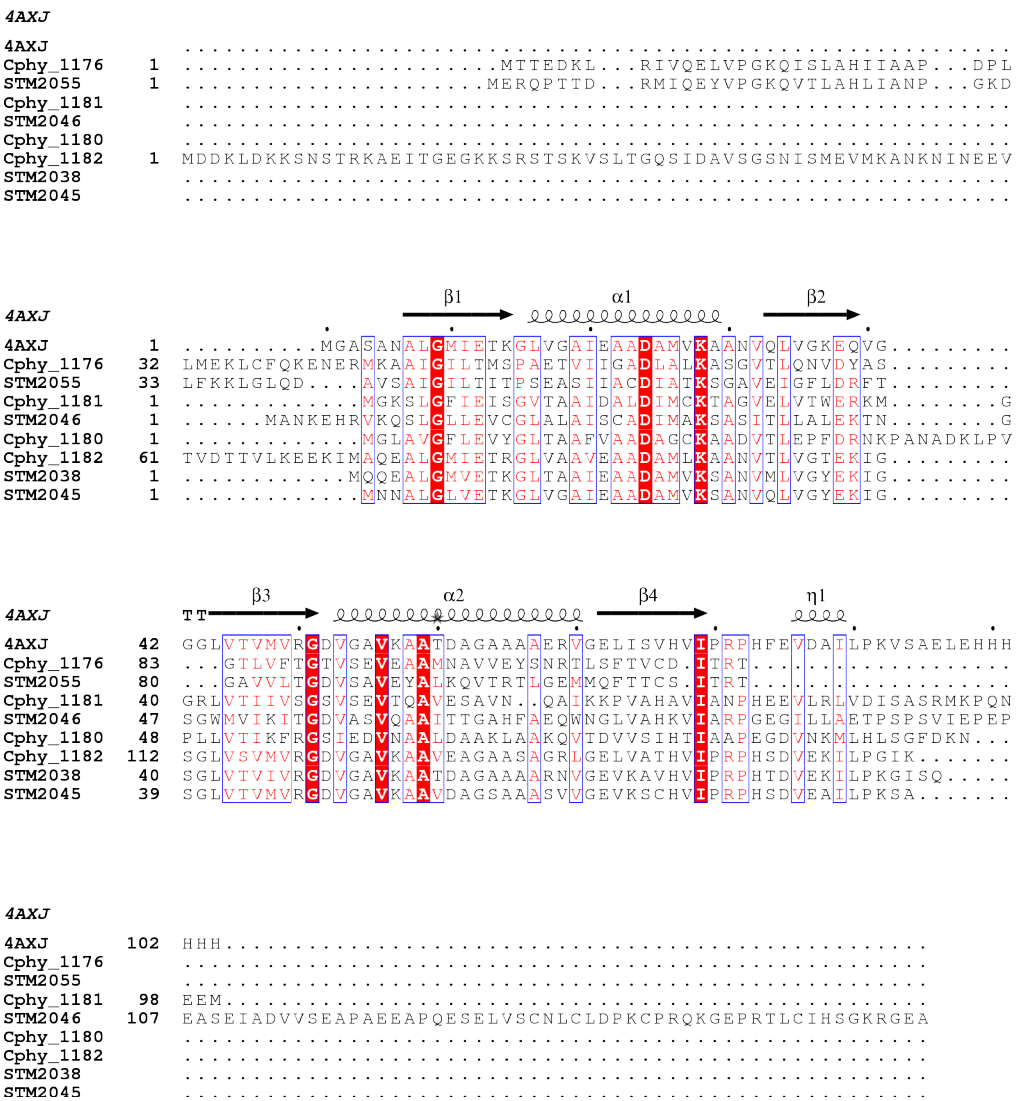


Figure 4.2 Multiple sequence alignment of the BMC-H type shell proteins used in this study, aligned against the EutM homologue from *C. difficile*.
The sequence alignment was generated using ClustalΩ (Sievers and Higgins, 2014)and displayed using ESPript (Gouet et al., 2003)4AXJ is the PDB code for the EutM homologue from C. difficile; the corresponding secondary structure elements from 4AXJ are shown above the alignment.

The β1-strand is relatively well conserved between the different BMC-H proteins, with a high level of conservation of the Gly8 residue. The Lys27 residue in CD1918 (α1) is shown to be involved in the crystal-packing region of the protein (Pitts et al., 2012) and again this residue is highly conserved between the BMC-H proteins shown in Figure 4.2.

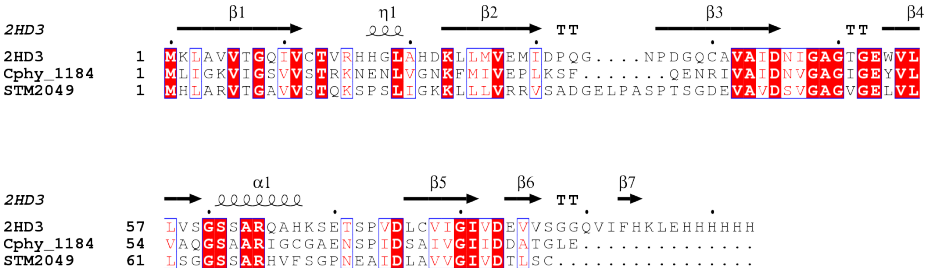


Figure 4.3 Multiple sequence alignment of the BMC-P type shell proteins used in this study, aligned against the EutN homologue from E. coli.
The sequence alignment was generated using ClustalΩ (Sievers and Higgins, 2014) and displayed using ESPript (Gouet et al., 2003). 2HD3 is the PDB code for the EutM homologue from *E. coli*; the corresponding secondary structure elements from 2HD3 are shown above the alignment.

The BMC-P-type proteins (Pfam 03319) were aligned against the EutN homologue from *E. coli* (Forouhar et al., 2007) (Figure 4.3). This shell protein type contains a five-stranded β-barrel, with an α-helix at the open end. There is a higher level of conservation between the two BMC-P-type proteins and the EutN homologue, than between the BMC-H-type and their EutM homologue.

The BMC-T-type protein alignment is shown in Figure 4.4. This particular homologue has been engineered to contain a 4Fe-4S cluster binding site (Aussignargues et al., 2016), although itself is a homologue of the PduT from *Citrobacter freundii* The STM2039 BMC-T protein is almost double the size of the other proteins in this alignment, however it still has alanine and glycine residue conservation in the β-1 strand. Overall these alignments show that there are some common conserved residues between the different shell protein types, mainly in the β1-strand.

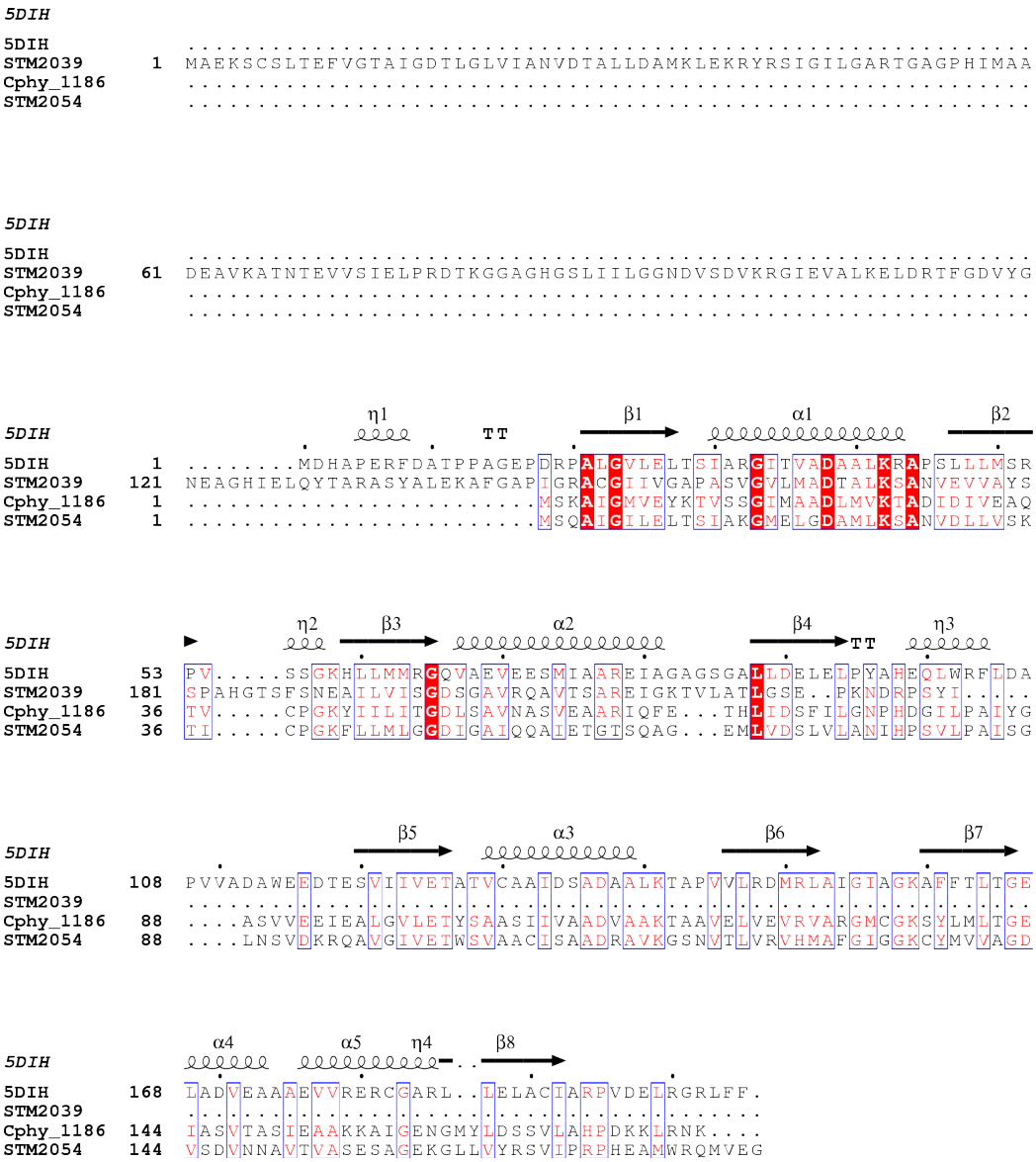


Figure 4.4 Multiple sequence alignment of the BMC-T-type shell proteins used in this study, aligned against the PduT homologue from *Citrobacter freundii*.
The sequence alignment was generated using ClustalΩ (Sievers and Higgins, 2014) and displayed using ESPript (Gouet et al., 2003). 5DIH is the PDB code for the PduT homologue from *Citrobacter freundii*; the secondary structure elements from 5DIH (Aussignargues et al., 2016) are shown above the alignment.

4.2.3 Establishing the MoClo DNA assembly pipeline for the production of synthetic BMCs

The Benchling software (available at Benchling.com) was used to design and test the assembly of all constructs in silico, including the initial primers used to clone the shell protein constructs (listed in Table 4.1), into the Level 0 plasmid. Both the PCR and digestion reactions were first checked in silico using the Benchling software, before

the primers were ordered. This step cuts down on cloning error due to weak/incorrect primer design and is an important part of the synthetic biology workflow. One of the BMC shell proteins, Cphy_1182 had an internal BpiI site, which needed to be mutated before this particular construct could be cloned into the Level 0 plasmid described in Materials and Methods

A design feature lacking from the MoClo toolkit parts was a series of common C-terminal protein tags, as well as a double stop codon to ensure ribosome detachment during protein translation and stop 'stop codon read-through'. Primers were designed to add His6, Strep, HA, and FLAG tags to the 5' end of the terminator constructs, such that in-frame codon reads would produce a tagged version of the protein coded in the upstream part. The PCR primers and detailed protocol for this method are described in Materials and Methods.

4.2.4 Rational design of shell protein constructs

Two interesting questions with respect to bacterial microcompartment shell formation are:

- Do the differences with respect to the BMC shell protein pores make chimeric synthetic microcompartments more effective when targeting specific synthetic enzymes and pathways to them?
- Does a shell protein's position on within the transcriptional unit affect the overall structure of the microcompartment shell?

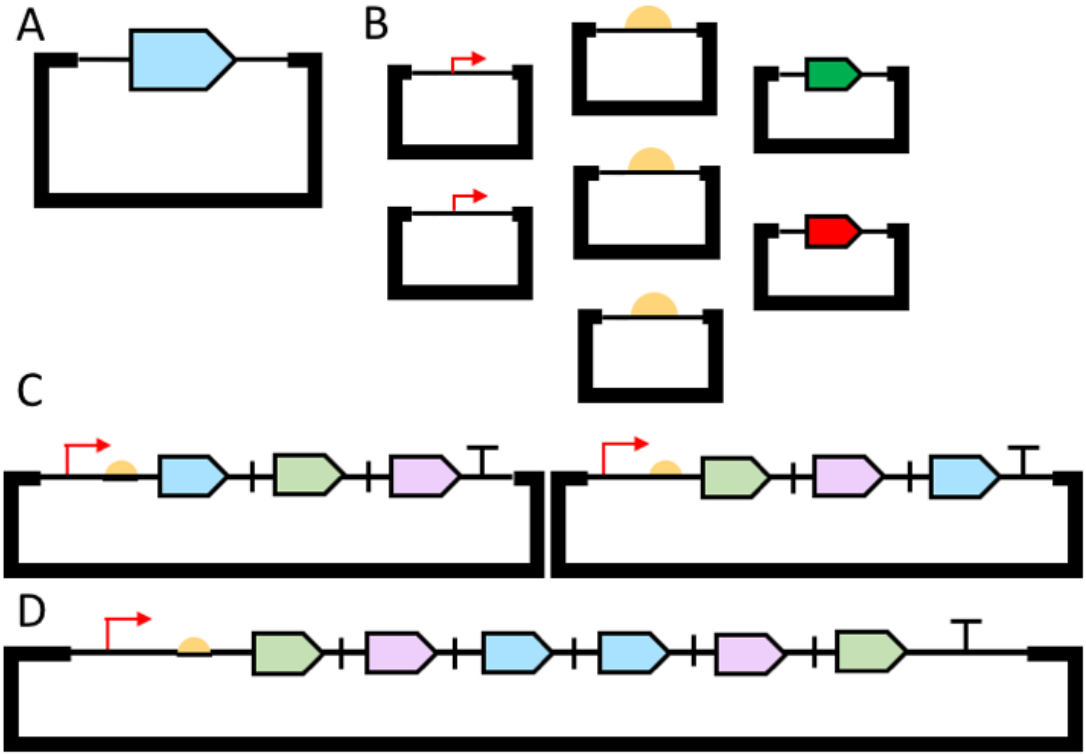


Figure 4.5 SBOL images of shell protein constructs, designed with Benchling software and generated using the modular cloning (MoClo) toolkit, in a one-pot assembly method.
A) A shell protein in the Level 0 ampicillin vector. B) A list of MoClo toolkit parts, including constitutive and controllable promoters (red arrows), high, medium and low strength RBSs (yellow semi-circles), fluorescent reporter genes (GFP and RFP in example above, green and red boxes respectively) and a series of terminators with different 3' overhangs for different levels of the system (black T). C) A level 1 transcriptional unit, including an RBS 'spacer' part (black vertical line) to ensure equal expression levels of individual shell proteins (purple, blue and light green boxes). D) A level 2 transcriptional unit made from a ligation of two level 1 transcriptional units. The different levels in the system allow larger constructs to be built, as a one-pot reaction success tends to drop with ligation of more than six molecules/one-pot reaction.

Despite it not being possible within the scope of this project to express any shell protein transcriptional units, constructs were designed as shown in Table 4.2.

	Cphy_1176	Cphy_1180	Cphy_1181	Cphy_1182	Cphy_1184	Cphy_1186	STM2038	STM2039	STM2045	STM2046	STM2049	STM2054	STM2055
Cphy_1176	P/R/S/R/S/T												
Cphy_1180	P/R/S/R/S/T	P/R/S/R/S/T											
Cphy_1181	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T										
Cphy_1182	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T									
Cphy_1184	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T								
Cphy_1186	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T							
STM2038	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T						
STM2039	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T					
STM2045	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T				
STM2046	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T			
STM2049	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T		
STM2054	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	
STM2055	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T	P/R/S/R/S/T

Table 4.2 Rational design of shell protein level 1 transcriptional units.
P=Promoter, R=RBS, S=Shell protein (colour differentiation to make design clearer), T=Terminator

Although the one-pot, Golden-Gate cloning is beneficial to generating this large number of constructs quickly, it is important to note that due to the design of the overhangs, it is not possible to control the order that the shell proteins are inserted into the transcriptional unit, or whether proteins will be duplicated.

4.2.5 Summary and discussion

Shell protein constructs were cloned into the MoClo toolkit Level 0 plasmids, with a view to building them into rationally designed transcriptional units, to study microcompartment shell formation. A series of terminator tags were cloned into the MoClo system, to make generating tagged constructs faster and more effective. Future work on this project would see a range of BMC shell protein cloned into the Toolkit, including Eut and Etu microcompartment proteins.

4.3 RBS + Cphy1178/STM2051 localisation sequence constructs

4.3.1 Introduction

For the assembly of BMCs, the question of how shell proteins are formed is also the question of how are the pathway enzymes targeted to the BMC? The two are not thought to be mutually exclusive, as a popular model for BMC assembly shows the core pathway enzymes recruiting the shell proteins via an encapsulation peptide, also known as a localisation sequence (Kerfeld et al., 2018). The localisation sequences from some of the key aldehyde dehydrogenases have been shown in Jacobson et al. (2015). An interesting paper from Fan et al. (2012) shows that the *Salmonella enterica* PduP enzyme (STM2051) localisation sequence interacts with the PduA shell protein via the Glu7, Ile10 and Leu14 residues, and that this interaction is critical fro the encapsulation of PduP into the BMC (Fan et al., 2012). Using the principal of enzyme encapsulation via the N-terminal localisation sequence, here the design of forward engineered ribosome binding site and localisation sequence-fused constructs from the PduP enzyme from *Salmonella enterica*, as well as the Cphy1178 aldehyde dehydrogenase is shown, along with generation of transcriptional units containing a

green fluorescent protein (GFP) or PhiLov reporter, to look at quantifying enzyme targeting to the bacterial microcompartment.

4.3.2 RBS + Cphy1178/STM2051 localisation sequence construct design

The RBS + localisation sequence constructs (RBS+loc) were designed to have different binding strengths (1X, 10X, 100X and 1000X) using the Howard Salis Ribosome Binding Site Calculator software. The constructs were ordered as complementary primers with suitable overhangs, annealed at 95ºC and subsequently cloned into the Level 0 B-C overhang plasmid (described in detail in Materials and Methods).

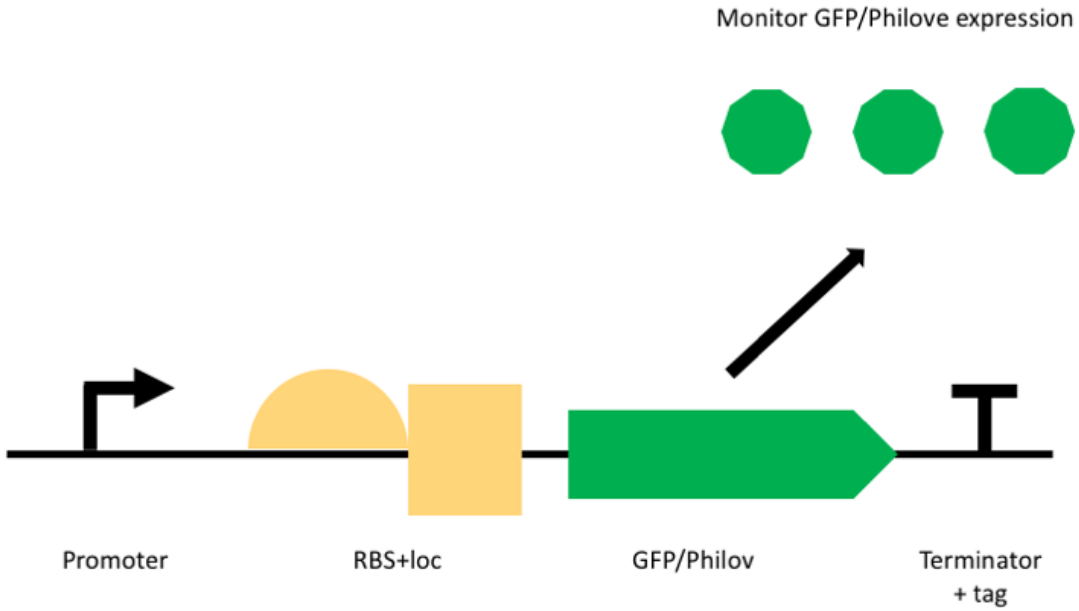


Figure 4.6 SBOL representation of RBS+loc construct transcriptional unit with fluorescent reporter gene.
The controllable/constitutive promoters and GFP are taken from the MoClo toolkit. Philov fluorescent reporter adapted for the MoClo system by Efrain Zarazua-Arvizu in the Marles- Wright Laboratory. Terminator and tag construct developed as described in section 4.2.3.

To correctly assay whether GFP had been encapsulated into the BMCs, and not just localised to the cytosol, two different ssrA-tags were cloned into the terminator plasmids from the Toolkit, to add a C-terminal degron to the fluorescent reporter

proteins. The ssrA –tags are recognised by the AAA+ ClpXP and ClpAP proteases and target the protein for degradation (Purcell et al., 2012).

Many Level 1 transcriptional units were generated for this project (Table 4.3). Preliminary data is shown for some of the constructs generated in section 4.1.3.3, however, further work on this project would refine the cell growth and GFP expression protocols, as described in (Wang et al., 2011).

TU	Promoter	RBS+loc	Reporter	Terminator	Expression system	Growth medium	Assay type
1	Controllable	All	GFP	Stop	*E. coli*	LB	Plate reader
2	Controllable	All	GFP	Stop	*E. coli*	M9 Minimal Media	Plate reader
3	Constitutive	All	GFP	Stop	*E. coli*	LB	Plate reader
4	Constitutive	All	GFP	Stop	*E. coli*	M9 Minimal Media	Plate reader
5	Constitutive	All	GFP	LAA degron tag	*E. coli*	M9 Minimal Media	Plate reader
6	Constitutive	All	GFP	LVA degron tag	*E. coli*	M9 Minimal Media	Plate reader
7	Constitutive	All	Philov	Stop	*E. coli*	M9 Minimal Media	Plate reader
8	Constitutive	All	GFP	LVA degron tag	*S. enterica*	M9 Minimal Media	Plate reader
9	Constitutive	All	GFP	LVA degron tag	*E. coli*	M9 Minimal Media	FACS

Table 4.3 Transcriptional units generated with the RBS+loc constructs.
Controllable promoter = R0010_AB (Toolkit part), Constitutive promoter = J23100_AB (Toolkit part), All = STM2051 and Cphy1178 (1X, 10X, 100X, 1000X) RBS+loc constructs, LVA and LAA degron tags design described in section 4.1.3.2. Expression system refers to the bacterial species with which the plasmid was transformed.

The main aim for this section of the project is to quantify in real-time, the number of GFP molecules encapsulate into the native *Salmonella enterica* BMCs using

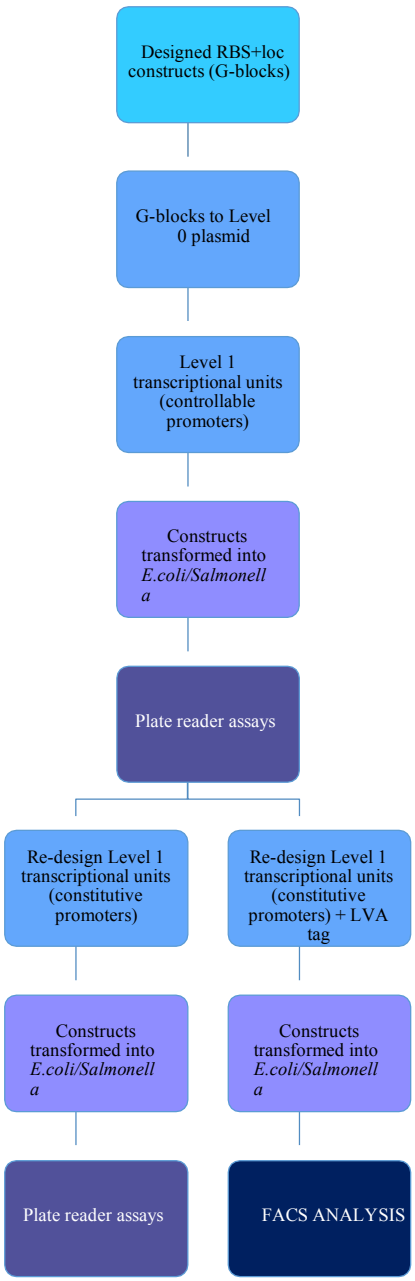


Figure 4.7 The design, build, test, learn, use synthetic biology workflow cycle as related to the RBS+loc sequence constructs.
Colours refer to Figure 4.1, and the different stages of the synthetic biology workflow. Ideally the construct design stage should drastically cut down bench-time, with efficient design protocols allowing for a quick and effective cloning process.

fluorescence activated cell sorting (FACS), and to see if rationally designed RBS strength ultimately has an effect on the amount of GFP localised to the microcompartments. However, initial parameters first have to be established to identify the most effective growth medium, growth conditions and a suitable

promoter. The workflow in Figure 4.7 shows the different stages of designing, building and testing these synthetic RBSs.

4.3.3 RBS + Cphy1178/STM2051 localisation sequence plate reader assays

The first transcriptional units designed for this section of the project used the controllable LacI promoter (R0010_AB) from the MoClo toolkit, the Cphy1178 RBS constructs (1X, 10X, 100X, 1000X), the GFP reporter gene, and the double-stop-modified Toolkit terminator. The constructs were transformed into the BL21 E. coli lac expression strain, and were cultured at 37ºC in LB medium. Figures 4.8, 4.9 and 4.10 show the results of these growth assays after 20 hours, with varying levels of IPTG.

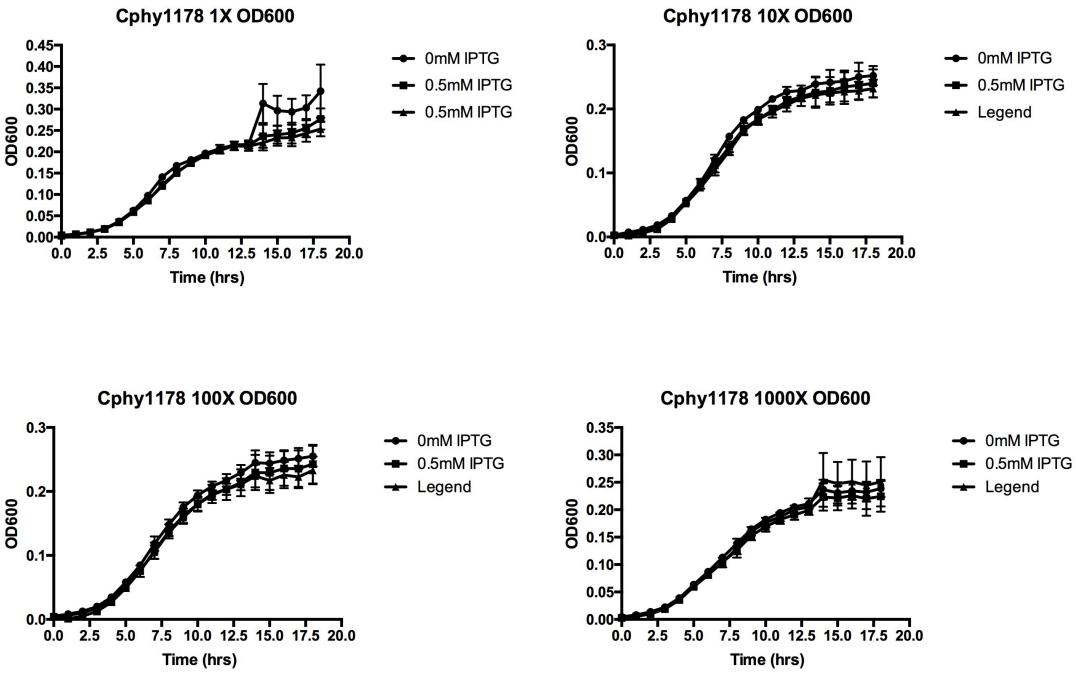


Figure 4.8 Time-course OD600 growth curves using construct TU1 from Table 4.3
E.coli cells were grown in LB broth at 37°C with shaking at 225rpm in a 96 well plate, with a starting OD600 0.01 and letting the cells grow to stationary phase. OD600 was monitored using a plate reader (SpectraMax M5). Cells were supplemented with different concentrations of IPTG (0.5 mM and 1 mM) to look at effect on GFP expression levels.

The OD600 measurements after 20 hours showed a standard sigmoidal growth curve typically expected for *E.coli* growth, with an apparent lag phase, log phase and stationary phase (Figure 4.8). The FL/OD600 curves show a dip at around the 5 hour mark for all constructs at the three different IPTG concentrations (Figure 4.10), with these values then rising to between 400 and 600 units after 20 hours. The GFP fluorescence was measured between wavelengths 485-538 nm. An initial reading of purified GFP protein was used to calibrate the fluorescence, however the values shown in Figure 4.9 are a lot lower than would typically be expected for this type of assay (Wang et al., 2011). This may be due to an error in construct design, although this is unlikely due to the prior in silico testing of the constructs. The GFP emission peak should be around 509 nM so is well within the range measured in the assay.

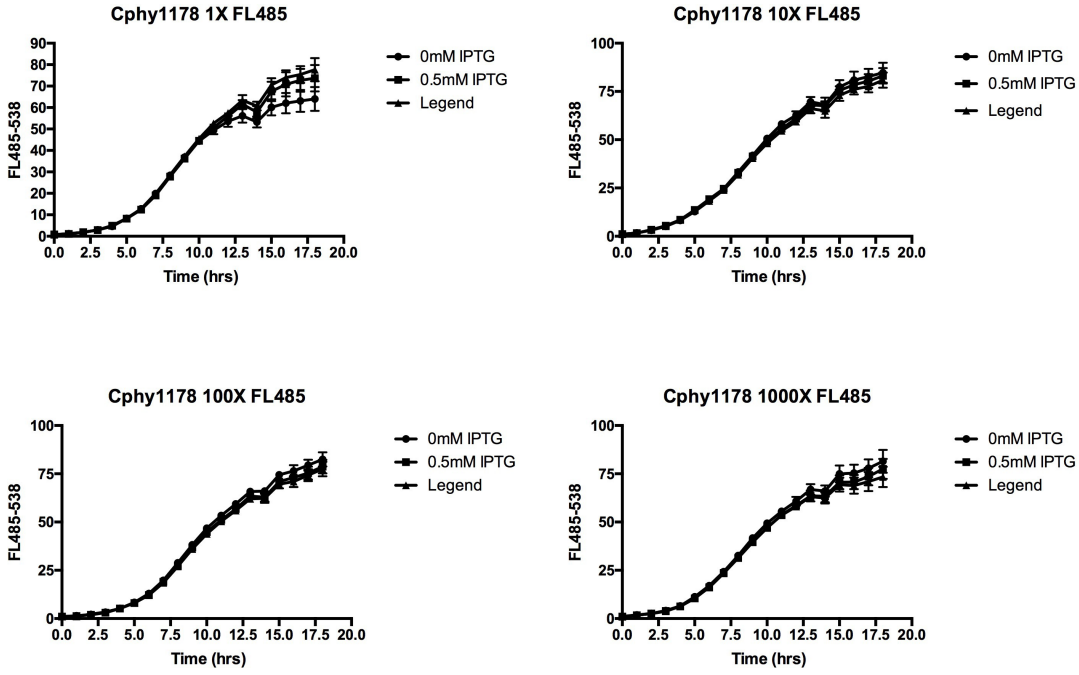


Figure 4.9 Time-course FL485-538 curves using construct TU1 from Table 4.3
E.coli cells were grown in LB broth at 37ºC with shaking at 225rpm in a 96 well plate, with a starting OD600 0.01 and letting the cells grow to stationary phase. FL485-538 was monitored using a plate reader (SpectraMaxM5). Cells were supplemented with different concentrations of IPTG (0.5 mM and 1 mM) to look at affect on GFP expression levels.

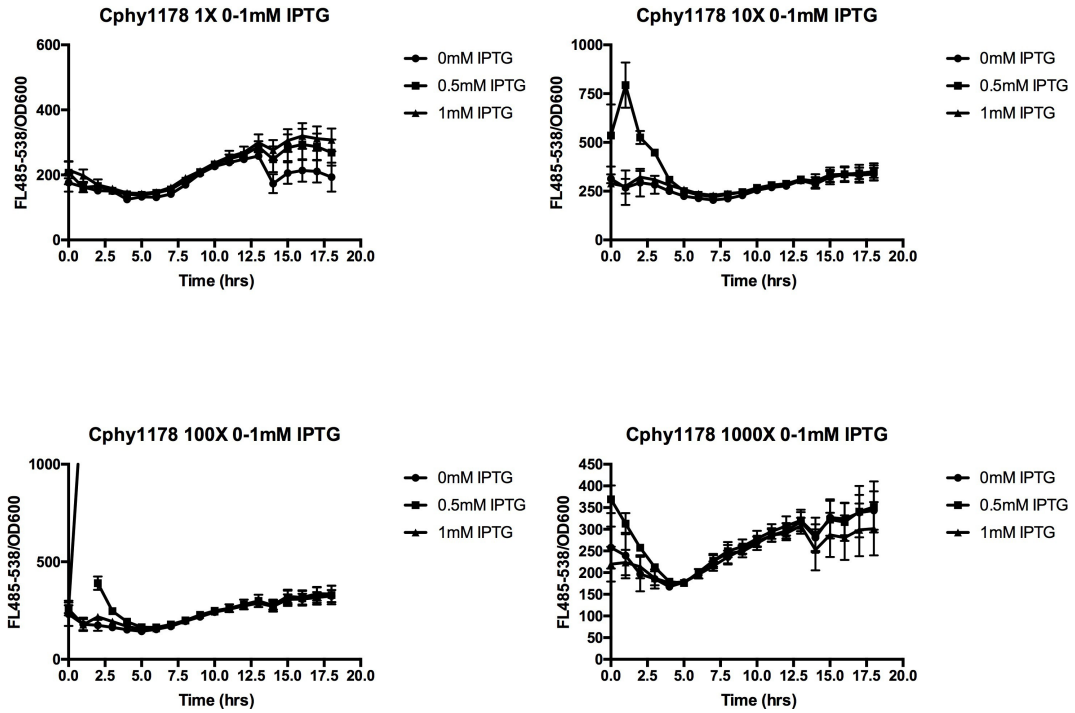


Figure 4.10 Time-course FL485-538/OD600 growth curves using construct TU1 from Table 4.3
E.coli cells were grown in LB broth at 37°C with shaking at 225rpm in a 96 well plate, with a starting OD600 0.01 and letting the cells grow to stationary phase. OD600 and FL485-538 was monitored using a plate reader SpectraMaxM5. Cells were supplemented with different concentrations of IPTG (0.5 mM and 1 mM) to look at affect on GFP expression levels.

Figure 4.11 shows the comparison of the different promoter strengths at 0.5 mM IPTG concentration. The values for FL/OD600 show that the 10X looks to have the highest GFP expression, although it isn't clear if the differences are statistically significant. A one-way ANOVA analysis was used to test if there is a significant difference between the different strength RBS constructs at each concentration of IPTG. The p values for all three IPTG concentrations were <0.00001, showing that there is a statistical significance between the constructs. Figure 4.12 illustrates the expression differences.

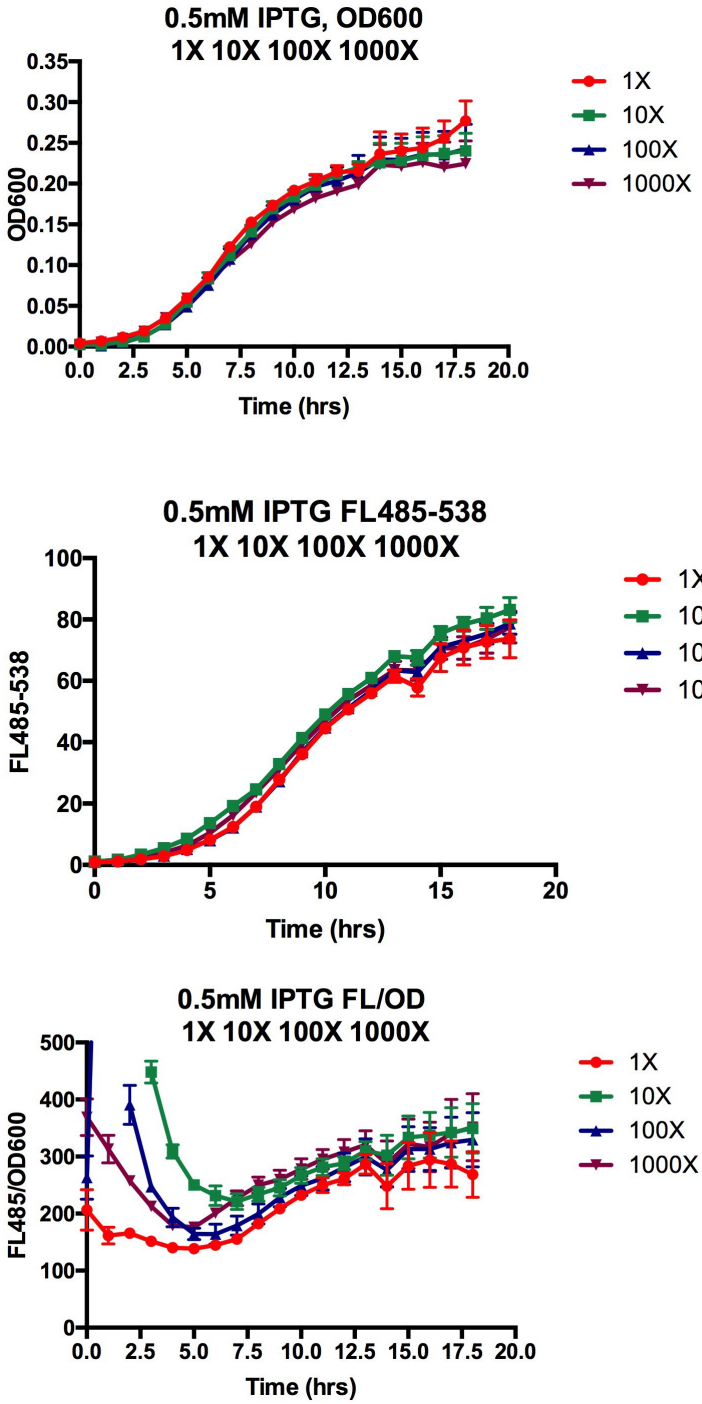


Figure 4.11 Time-course OD600, FL485-538 and FL485-538/OD600 growth curves using construct TU1 from Table 4.3 measured at IPTG concentration of 0.5 mM.
E.coli cells were grown in LB broth at 37°C with shaking at 225rpm in a 96 well plate, with a starting OD600 0.01 and letting the cells grow to stationary phase. OD600 and FL485-538 was monitored using a plate reader SpectraMaxM5. Cells were supplemented with 0.5 mM IPTG to look at affect on GFP expression levels.

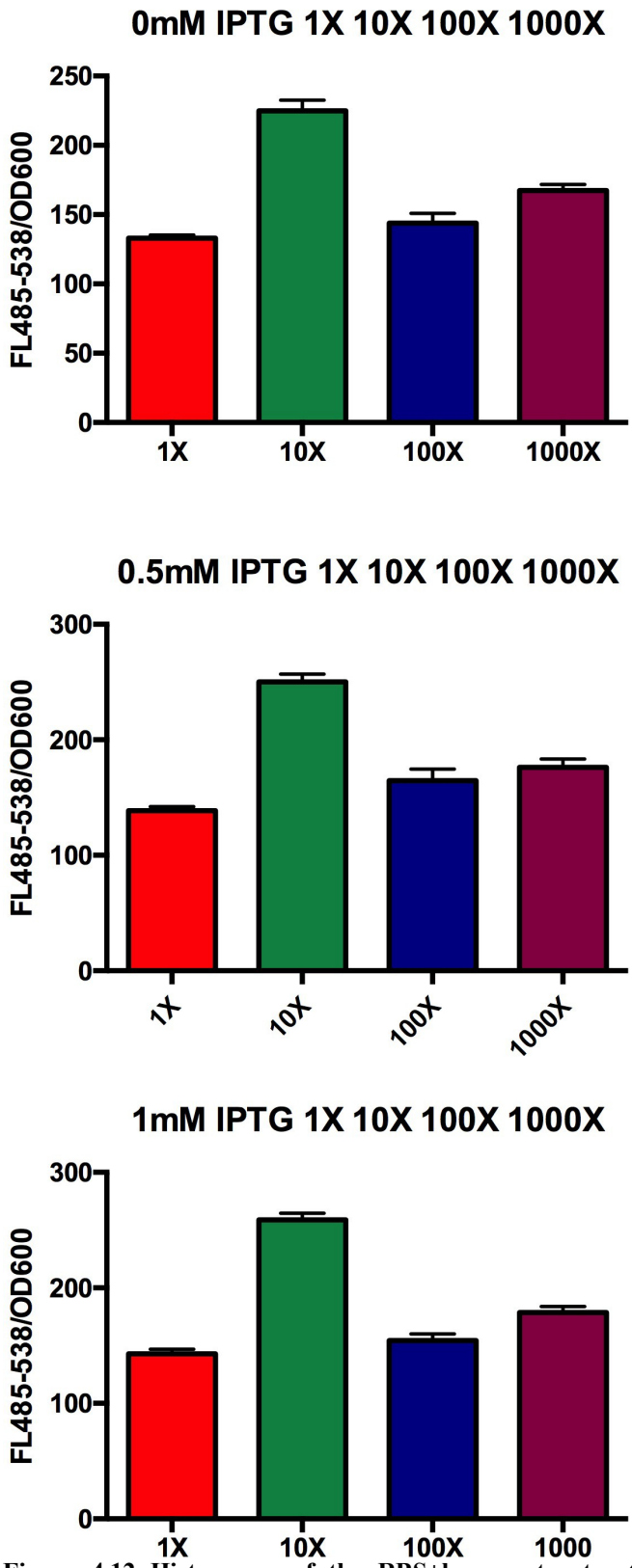


Figure 4.12 Histograms of the RBS+loc constructs at 0 mM, 0.5 mM and 1 mM IPTG concentration.

The histograms show that the highest FL/OD600 value is for the 10X RBS construct, at all concentrations of IPTG tested. The results of a one-way ANOVA showed that there is statistical significance between the each of the RBS constructs

4.3.4 RBS + Cphy1178/STM2051 localisation sequence plate FACS analysis

For the preliminary FACS analysis, the TU4 construct (Table 4.3) was used. The change to the constitutive promoter meant that an overall basal expression level was established, without the need for induction with IPTG. The constructs had the STOP terminator, were transformed into BL21 E. coli cells and were grown in M9 minimal medium. Figure 4.13 shows the results of the FACS sorting analysis, under the conditions specified in Materials and Methods. The histograms of Cphy1178 constructs for both the plate reader assays and FACS analysis (Figures 4.12 and 4.13) show the highest FL/OD600 values and geometric mean values for the Cphy1178 10X RBS+loc construct. The geometric mean is used for FACS, as it useful to compare the single 'figure of merit' - in this case GFP fluorescence count- rather than when each item has multiple properties that have different numeric ranges.

4.3.5 Summary and discussion

These preliminary results cannot be directly compared as the growth medium differs between plate reader and FACS experiments (LB and M9 minimal medium respectively). For the STM2051 RBS+loc constructs, the mean increases with the strength of the designed RBSs. The geometric mean values for the STM2051 1000X are almost 6 times higher than for the Cphy1178 1000X constructs, although much more work is needed to conclusively prove a quantifiable difference between RBS+loc construct strength discussed in detail in Chapter 5.

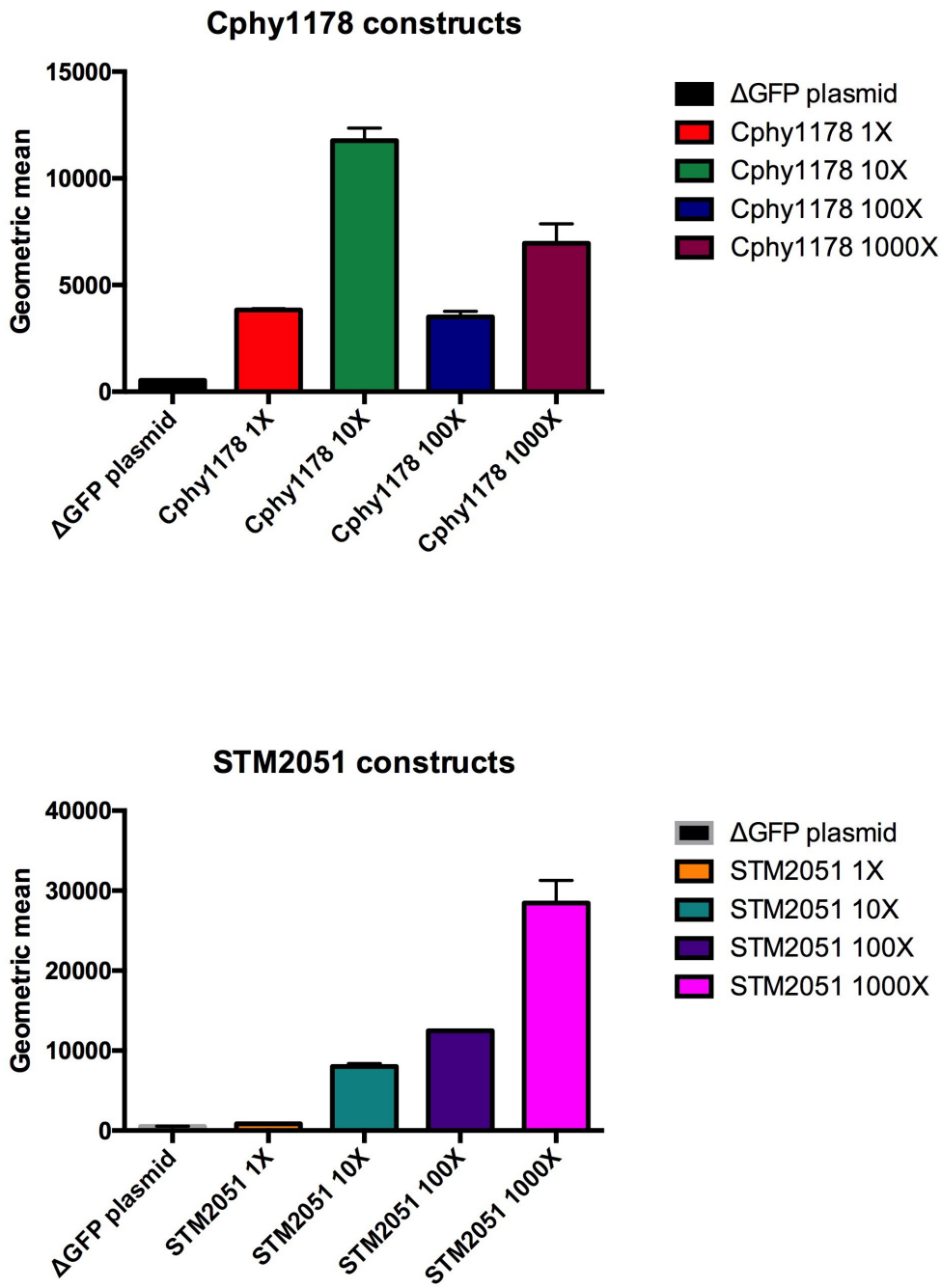


Figure 4.13 Histograms of the geometric means from FACS sorting analysis for Cphy1178 and STM2051 RBS+loc constructs

4.4 EutM shell proteins

4.4.1 Introduction

Different homologues of the common EutM protein show different charge states across the surface of the protein (Held et al., 2013; Pitts et al., 2012), with EutM having a strongly negatively charged pore (Takenoya et al., 2010) and the homologue CD1918 from *Clostridium difficile* having a more even charge across the surface (Pitts et al., 2012). This suggests that there is a difference in charge states across the surface of the protein, not only between EutM domains from different species, but also between the same species.

Overexpression of EutM protein tends to lead to aberrant growth in cells. The aim of rationally expressing BMC shell proteins is to use them as scaffolding systems (Noël et al., 2016; Quin et al., 2017), although more needs to be known about EutM protein surface charge and pore before this would be possible.

4.4.2 pET28a-GG-RFP plasmid for one-pot cloning of ORFs for overexpression

The commonly used pET28a vector is useful for cloning as it contains the T7 promoter system, for in-frame protein production under IPTG control, as well as an inbuilt His6 tag on the 3' end of the multiple cloning site (MCS). However a problem with standard cloning is that it can be time consuming, as two separate enzymes are generally needed to insert the PCR product into the vector. By modifying the pET28a vector with the C- and D- defined overhangs, it is possible to quickly clone into a strong protein expression vector, using the same one-pot reaction mechanism as with the MoClo system.

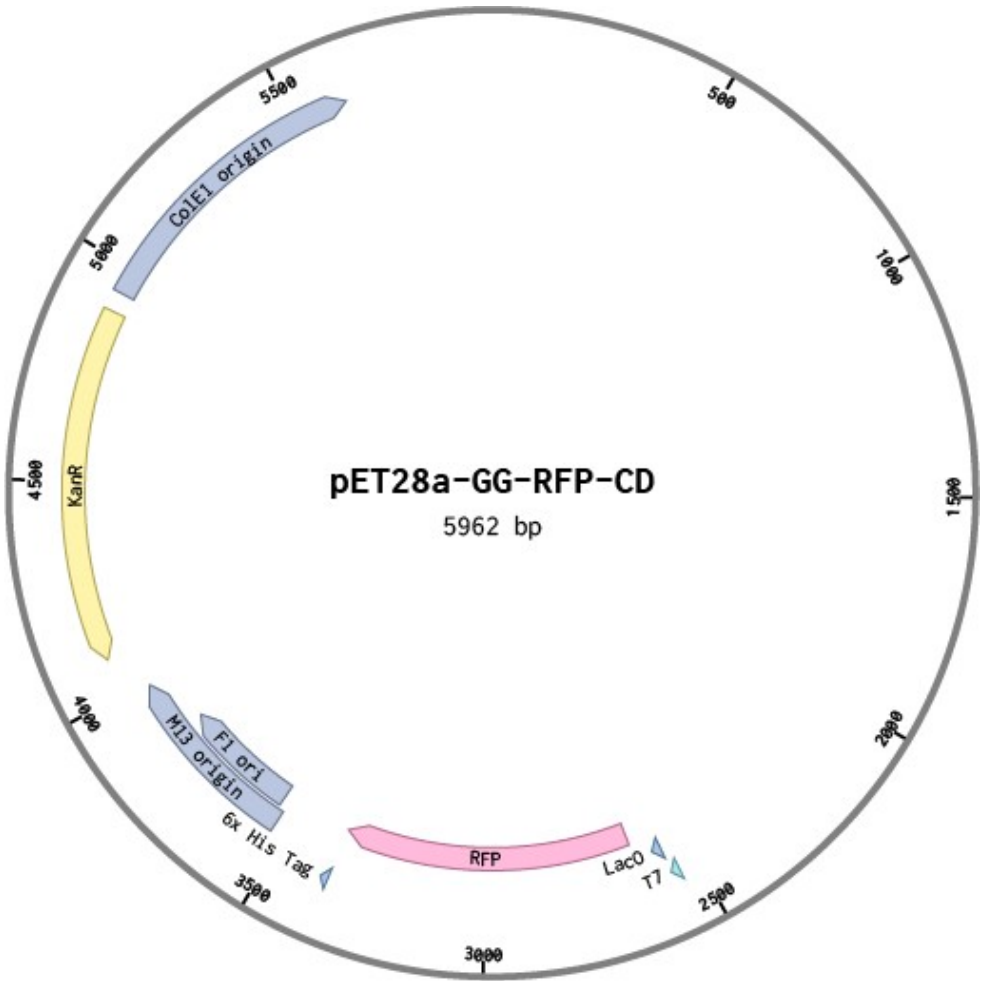


Figure 4.14 Plasmid map of the pET28a vector, engineered into a MoClo compatible protein expression vector
The RFP fluorescent reporter is knocked out when a construct is successfully inserted between the C – and D- overhangs, so if plated onto LB agar supplemented with IPTG, it is possible to differentiate which colonies have cloned in the gene of interest.

4.4.3 Construct expression and purification by ammonium sulfhate precipitation

EutM constructs from Des*ulfotalea psychrophila*, *Psychromonas hadali*, *Salmonella enterica* and *Thauera linaloolentis* were received from Maureen Quin in the Schmidt-Dannert lab (University of Minnesota), and were cloned into the golden- gate pET28a plasmid, both with and without a stop codon, to generate untagged and His6-tagged constructs. All four of the constructs expressed well under standard over-expression conditions (growth to OD600 0.6, induction with 1 mM IPTG, 4 hours growth at

37 ºC in 1 litre cultures and so were subject to purification. Figures 4.15 and 4.16 show the anion exchange purification of the untagged *D. psychrophilia* (Dp) and *P. hadalis* (Ph) constructs. The majority of the protein was found in the flow-through fraction for both constructs, and formed a cloudy/iridescent lysate, which was found to be soluble protein, as centrifugation of the samples did not form a pellet. It was suggested by our collaborators that the EutM protein was forming sheet-like structures in solution, so it was decided to further purify the sample for further analysis by atomic force microscopy (AFM) as carboxysome shell proteins had been previously studied to great success by this method (Erbilgin et al., 2016).

A second purification step of size-exclusion chromotagraphy did not give sufficient amounts of protein for downstream analysis by AFM, so purification by ammonium sulfate precipitation was introduced. Figure 4.17 shows that most of the Dp construct was recovered at 10% w/v and 20% w/v ammonium sulfate concentration; although there is discernible protein for the 30%, 50% and 80% fractions. Ph shows faint protein bands in the 10%, 20% and 30% fractions, however the majority of the protein was purified from the 50% and 80% fractions.

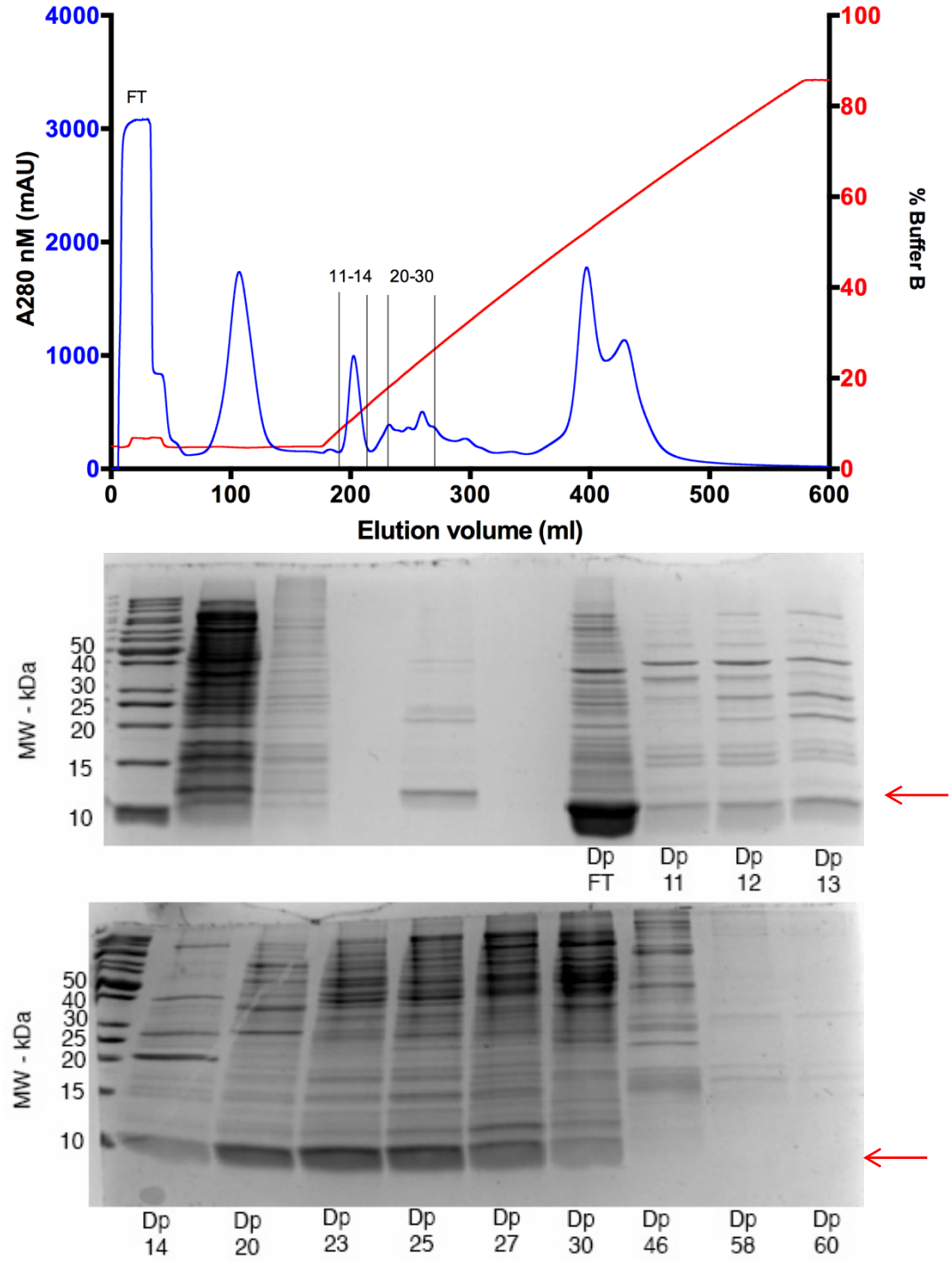


Figure 4.15 Anion exchange purification of recombinant EutM from *Desulfotalea psychrophila.*

Fraction FT corresponds to the flow-through. Fractions were eluted by a linear gradient over 20 column volumes from 0–1 M NaCl as fractions from FT-60. Fractions from FT to 60 were analysed by 15% w/v SDS-PAGE, with the majority of the protein (size denoted by red arrow, 10 kDa in the flow-through fraction.

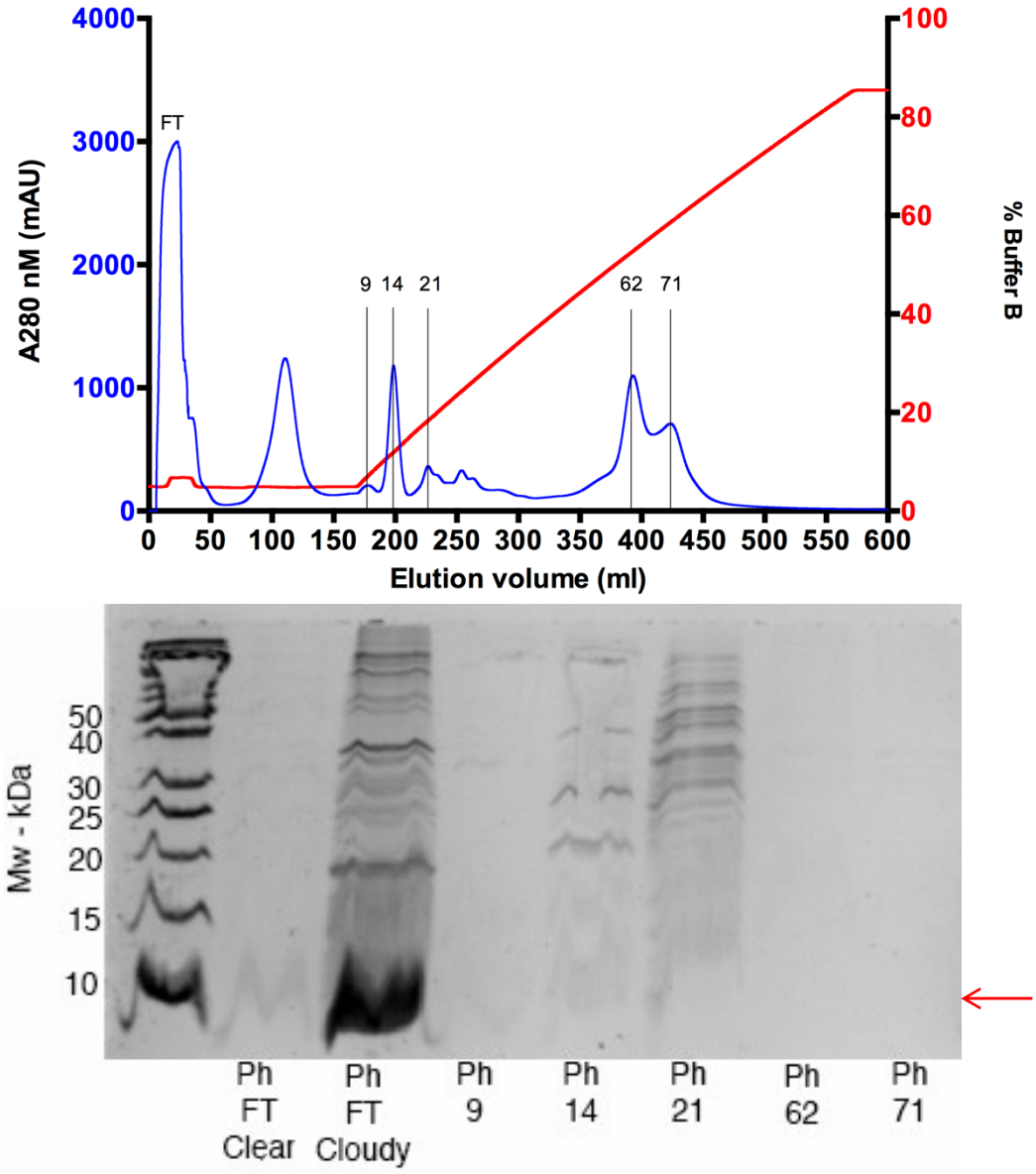


Figure 4.16 Anion exchange purification of recombinant EutM from *Psychromonas hadalis*
Purification from a 12 ml Q-sepharose column Fraction FT corresponds to the flow-through. Fractions were eluted by a linear gradient over 20 column volumes from 0–1 M NaCl as fractions from FT-71. Fractions from FT to 71 were analysed by 15% w/v SDS-PAGE, with the majority of the protein (size denoted by red arrow, 10 kDa) in the flow-through fraction.

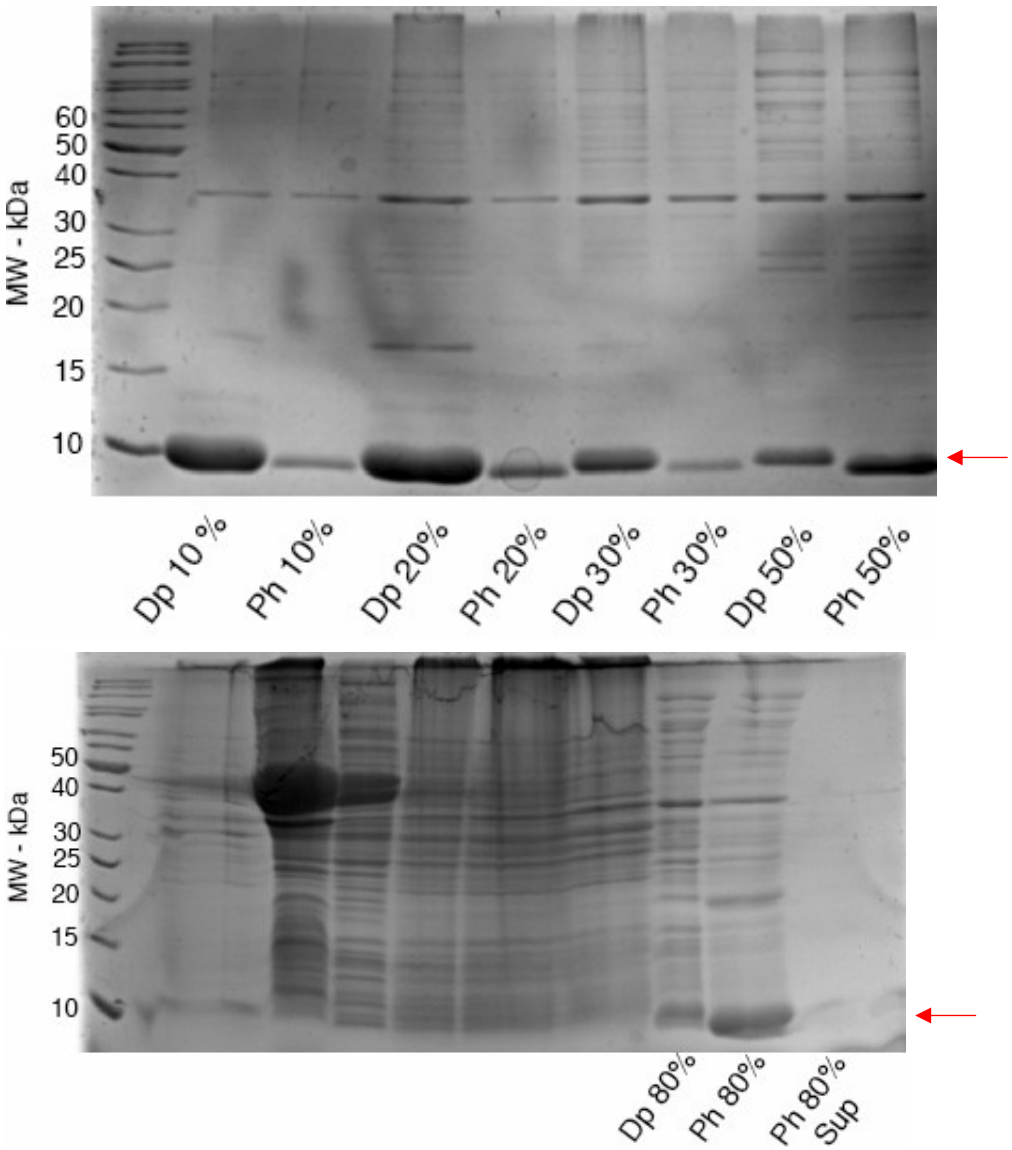


Figure 4.17 Ammonium sulfate precipitation of recombinant EutM from *Desulfotalea psychrophila* (Dp) and *Psychromonas hadali* (Ph).
Flow-through fractions of Dp and Ph from anion exchange were subject to ammonium sulfate precipitation. Proteins were sequentially precipitated from 50 ml of crude extract by stepwise addition of solid ammonium sulfate with stirring at a certain degree of saturation. The pellet obtained after each centrifugation step was resuspended in max. 10 ml anion exchange buffer. Steps were repeated for 10, 20, 30, 50 (top gel) and 80% (bottom gel) ammonium sulfate and fractions were analysed by 15% w/v SDS-PAGE. Size denoted by red arrow, 10 kDa.

The second stage of purification did not produce a clear lysate, although centrifugation of the lysate suggested that the protein was still in solution as no pellet was formed. It was decided that thin-section electron microscopy (TEM), would be the next step to show the EutM proteins in vivo and to ascertain their structure inside the cell.

4.4.4 Thin-section electron microscopy of EutM shell proteins

Although only the Dp and Ph EutM constructs were expressed and purified on a larger scale due to time constraints, the untagged and His6-tagged variants of EutM Dp, Ph and Stm and the untagged variant of Tl were subjected to TEM. It was not possible to produce the His6-tagged Tl construct in the timeframe of the project, so only the untagged version was available for TEM. The results of this experiment were interesting; the tagged and untagged versions of the constructs all showed normal E. coli cell growth in liquid culture; however, the morphology of the cells was highly distorted, with the cells elongated up to double their normal size of 1-2 μm (Figure 4.18 (C) and Figure 4.20 (A). It appears that the *E. coli* cell is elongated due to its inability to separate into two daughter cells due to the sheets of EutM protein running through the centre of the cell and interfering with the correct formation of a competent division septum. At first glance there is little difference between the untagged and His6-tagged version of the EutM constructs, with both variants showing this distorted morphology across the four different protein species (Figures 4.18, 4.19, 4.20 and 4.21). The transverse views of each construct (Figures 4.18 (B) and (D), Figure 4.19(B) and (D), Figure 4.20 (B) and (C) and Figure 4.21 (B) and (C)), show a tight spiral-like arrangement of EutM protein spanning the entire diameter of the cell. In Figure 5.9, it appears that there is a slight difference in the diameter of the protein spiral, with the un-tagged Dp construct (Figure 4.18 (B)) appearing to be slightly smaller and more tightly curved than in the His6-tagged construct (Figure 4.18 (D)). However, it is difficult to prove the same effect between the other three species constructs (Ph, Stm and Tl). A reason for this could be that the His6 tag interferes with the EutM protein spiral, leading to a less-tightly packed protein conformation.

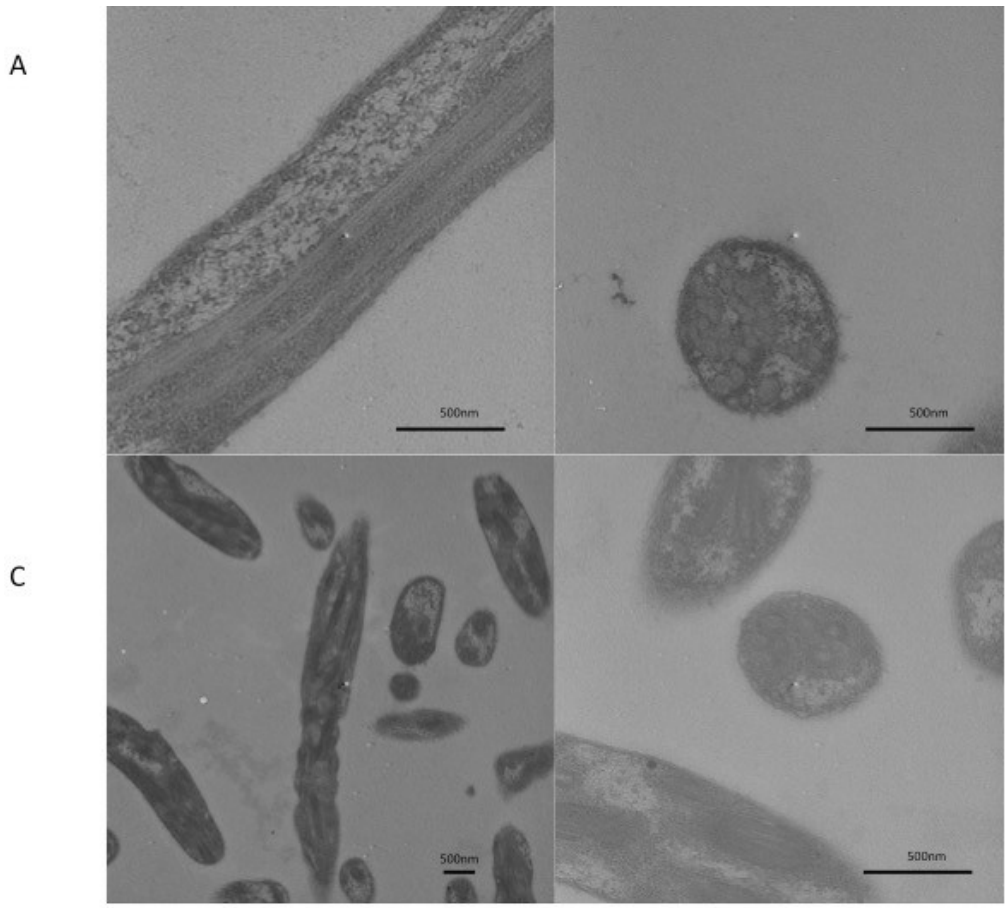


Figure 4.18 Thin-section transmission electron micrograph images of overexpression of recombinant EutM protein from *Desulfotalea psychrophila* in *E. coli* cells
Images taken with the Philips CM100 TEM at Newcastle University. A) Longitudinal view of *E. coli* showing overexpression of untagged EutM from *D. psychrophila*. Cell shows distorted and elongated morphology, image taken at 46,000 x magnification. B) Transverse view of *E. coli* showing overexpression of untagged EutM from *D. psychrophila*. Cell shows spiral arrangement of EutM protein clusters, spanning the entire cell. Image taken at 46,000 x magnification. C) Longitudinal and transverse views of multiple E. coli cells showing overexpression of His6 tagged EutM from *D. psychrophila*. Image taken at 13,500 x magnification. D) Longitudinal and transverse views of multiple E. coli cells showing overexpression of His6 tagged EutM from *D. psychrophila*. Image taken at 46,000 x magnification. Both tagged and untagged versions of the EutM protein seem to show similar morphology within the cell, however from the transverse views it appears that the untagged proteins show a tighter spiral than the His6 tagged version.

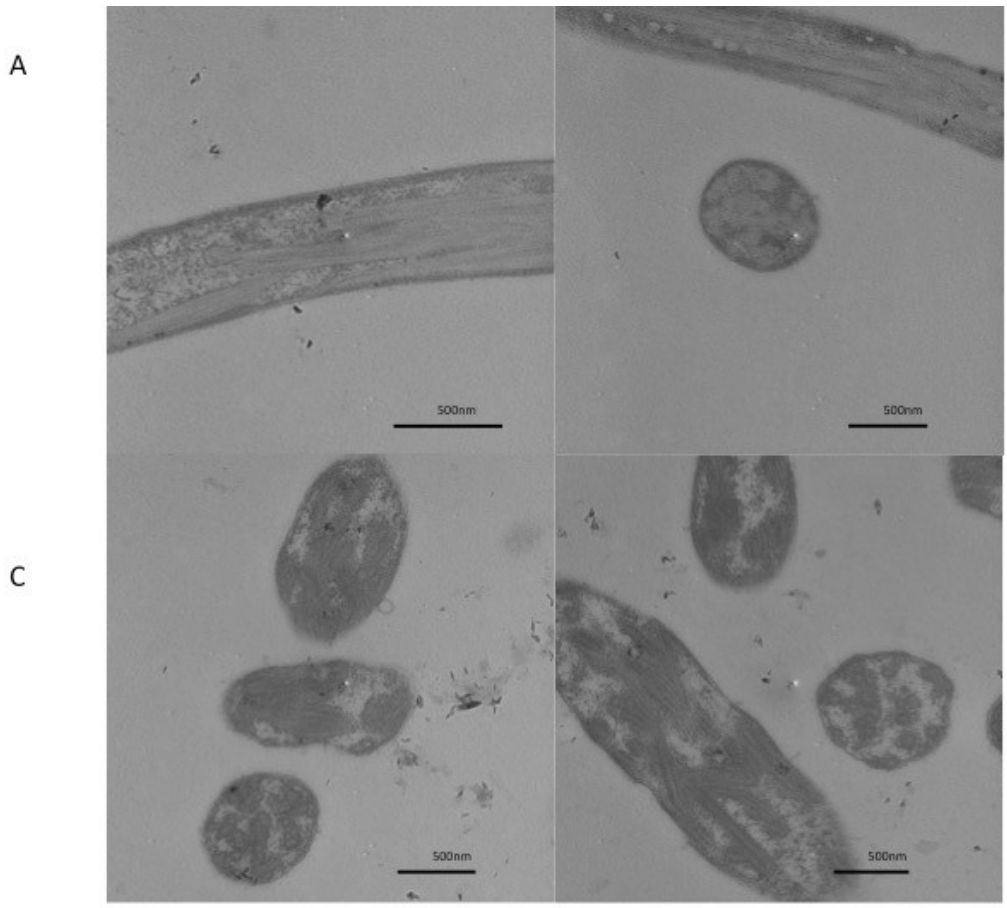


Figure 4.19 Thin-section transmission electron micrograph images of overexpression of recombinant EutM protein from Psychromonas hadalis in E. coli cells

Images taken with the Philips CM100 TEM at Newcastle University. A) Longitudinal view of *E.coli* showing overexpression of untagged EutM from *P. hadalis*. Cell shows distorted and elongated morphology. Image taken at 34,000 x magnification. B) Transverse view of E. coli showing overexpression of untagged EutM from *P. hadalis*. Cell shows spiral arrangement of EutM protein clusters, spanning the entire cell. Image taken at 34,000 x magnification. C) Transverse views of multiple E. coli cells showing overexpression of His6 tagged EutM from *P. hadalis*, image taken at 34,000 x magnification. D) Longitudinal and transverse views of multiple E. coli cells showing overexpression of His6 tagged EutM from P. hadalis. Image taken at 46,000 x magnification. Both the longitudinal and transverse views show overall similar size and shape of protein structures for tagged and untagged EutM.

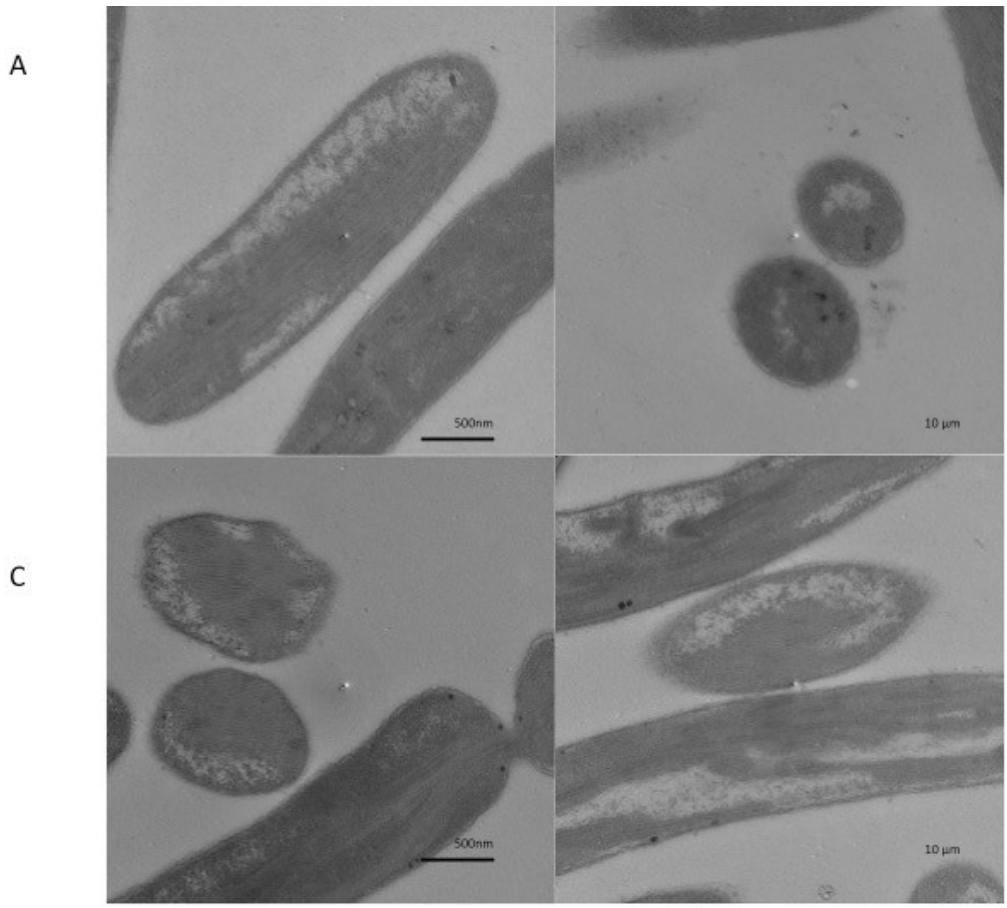


Figure 4.20 Thin-section transmission electron micrograph images of overexpression of recombinant EutM protein from *Salmonella enterica* in *E. coli* cells

Images taken with the Philips CM100 TEM at Newcastle University. A) Longitudinal view of *E. coli* showing overexpression of untagged EutM from *S. enterica*. Cell shows distorted and elongated morphology. Image taken at 34,000 x magnification. B) Transverse view of *E. coli* showing overexpression of untagged EutM from *S. enterica.* Cell shows faint dots, relating to the overexpressed EutM protein. Image taken at 11,000 x magnification. C) Transverse views of multiple E. coli cells showing overexpression of His6 tagged EutM from *S. enterica*. Image taken at 34,000 x magnification. The EutM protein appears to be forming sheets, rather than spiral-like arrangements. D) Longitudinal and transverse views of multiple E. coli cells showing overexpression of His6 tagged EutM from S. enterica. Image taken at 11,000 x magnification. The comparison between tagged and untagged EutM is not as clear from these images, although from the longitudinal views in A) and C), the protein arrangement looks to be similar.

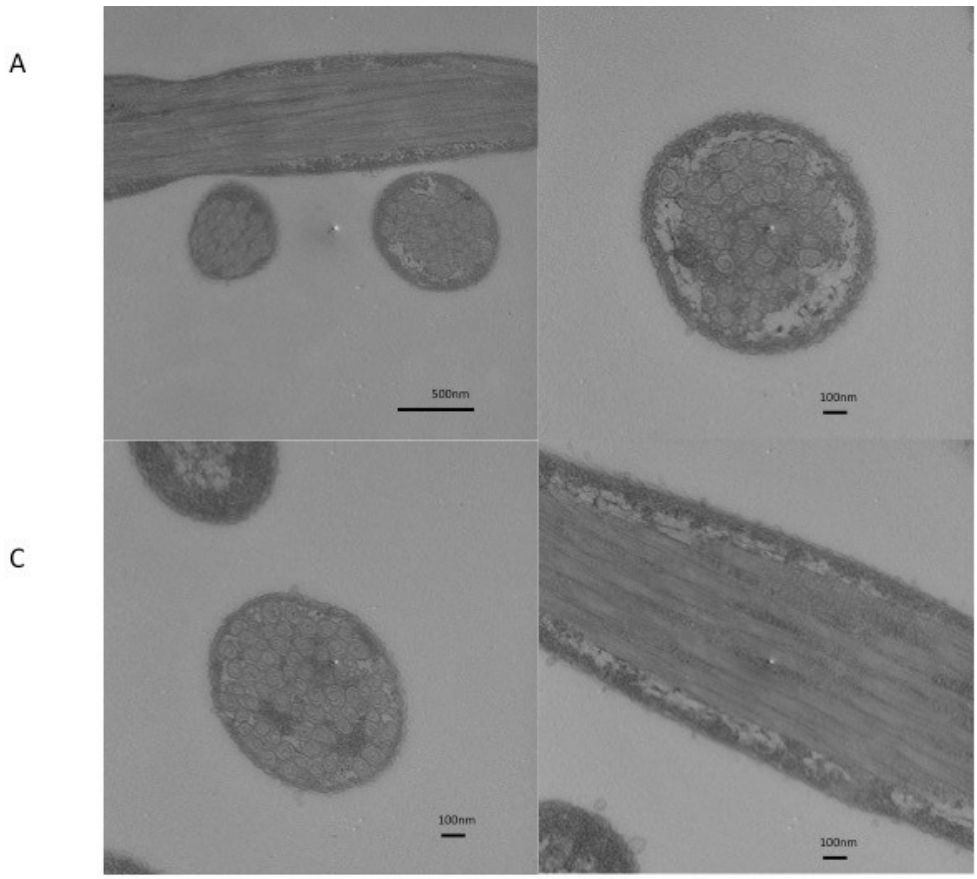


Figure 4.21 Thin-section transmission electron micrograph images of overexpression of recombinant EutM protein from *Thauera linaloolentis* in E. coli cells

Images taken with the Philips CM100 TEM at Newcastle University. A) Longitudinal and transverse views of *E. coli* showing overexpression of untagged EutM from *T. linaloolentis*. Cells shows distorted and elongated morphology. Image taken at 34,000 x magnification. B) and C) Transverse view of E. coli showing overexpression of untagged EutM from *T. linaloolentis*. Both images show multiple, spiral-like arrangements of EutM proteins, averaging around 40 spirals/cell. Images both taken at 64,000 x magnification D) Longitudinal view of E. coli showing overexpression of untagged EutM from *T. linaloolentis*. Image taken at 64,000 x magnification. The EutM protein structures are shown to run through the entire length of the cell. The His6 tagged version of the *T. linaloolentis* EutM was unable to be cloned.

4.4.5 Summary and discussion

It has been shown that the EutM protein has a concave and convex side, with the concave side holding the N- and C- terminal of the protein (Held et al., 2013); these constructs have been generated with a C-terminal His6 tag and so the tag is likely to also reside in the concave side of the protein. Different homologues of the EutM protein show different charge states across the surface (Held et al., 2013; Pitts et al., 2012), with EutM having a strongly negatively charged pore (Takenoya et al., 2010) and the homologue CD1918 from Clostridium difficile having a more even charge across the surface (Pitts et al., 2012)This suggests that there is a difference in charge states across the surface of the protein even between the same protein from different species. Although the TEM images from the four different EutM examples shown here do not conclusively show that there is a difference between the protein packing, the fact that overexpression of the protein forms spirals in itself suggests that the concave and convex sides to the protein have differing charge states, and this could affect the packing of the protein within the spirals.

5 Discussion

5.1 Expression and purification of acylating aldehyde dehydrogenase enzymes

The truncated versions of three putative aldehyde dehydrogenase enzymes from *Clostridium phytofermentans* were expressed as recombinant proteins in *E. coli* BL21(DE3) cells. The enzymes were purified in two stages, first by ion-exchange chromatography, over a 12 ml Q-sepharose column and further by size-exclusion chromatography, which yielded sufficient pure protein for further analysis and experiments.

The enzymes were then used for further experiments to characterise them biochemically and structurally; including the aldehyde dehydrogenase assays and structure determination by X-ray crystallography. From the well-expressed Cphy1178_20 construct, two variants were proposed, C269A and H387A; these were cloned, expressed and purified. These variants were then tested for specific activity along with the wild-type protein and were also subject to further structural analysis.

5.1.2 Activity of aldehyde dehydrogenase Cphy1178 against short chain fatty-aldehydes

Cphy1178 was tested for activity against the C2-C6 short-chain fatty aldehydes; it displayed the highest turnover rate for propionaldehyde, confirming that this enzyme is a propionaldehyde dehydrogenase as inferred from the BMC encoded pathway in the *C. phytofermentans* fucose/rhamnose BMC locus. The active site variants C269A and H387 showed no activity against the range of aldehydes tested. Recent papers from (Zarzycki et al., 2017)and (Becher et al., 2018) confirm the catalytic mechanism proposed, with characterisation of a propionaldehyde dehydrogenase homologue from

Rhodopseudomonas palustris, that has the highest kcat/KM values against the propionaldehyde substrate. Substrate inhibition was also noted at higher aldehyde concentrations, which is consistent with a yeast aldehyde dehydrogenase enzyme and the data presented in this thesis. The reason suggested for this inhibition at higher concentrations is that a dead-end enzyme-NADH- aldehyde complex may be formed (Wang et al., 2009).

5.1.3 Crystal structures of aldehyde dehydrogenase Cphy1178 with NAD^+ and Co-enzyme A in the nucleotide binding domain

The Cphy1178 aldehyde dehydrogenase was crystallised with the NAD+ cofactor in the cofactor-binding pocket; similarly the Cys269A variant was crystallised with Co-enzyme A in the same binding pocket although, in this complex the adenine ring lies in the opposite conformation. The proposed mechanism of action for this enzyme is that the H387 residue deprotonates the catalytic cysteine to allow nucleophilic attack on the substrate to form the tetrahedral acyl-enzyme intermediate.

Following hydride transfer to the NAD^+ cofactor, the NADH leaves the binding pocket, followed by entry of the CoA cofactor. The E357 residue is proposed to deprotonate the CoA, which then attacks the acyl-enzyme intermediate to form the thioester product and free enzyme. The extra proton on E357 is suggested to be transferred to water. However, the paper by Zarzycki et al. (2017), suggests a slightly different mechanism for the *R. palustris* homologue. They state that the acylated cysteine residue adopts a different rotamer conformation, which makes it inaccessible to the conserved glutamic acid residue. They propose that the H449 residue is involved in the deprotonation of CoA via two water molecules.

Before the publication of (Zarzycki et al., 2017), an E357A variant had been proposed to confirm that this residue is involved in the deprotonation of the CoA molecule. The crystal structure of this variant shows a CoA molecule in the expected conformation in the cofactor binding pocket. However, these were apo-crystals and were not exposed to CoA experimentally. A suggestion for this observation is that the CoA has become 'trapped' in the binding pocket, because the proton had not been abstracted. The two-models are not particularly mutually-exclusive in this regard and it is possible that both the water molecules and E357 residue have a part to play in the deprotonation of CoA.

5.2 Building synthetic microcompartments

The MoClo toolkit, developed by the Densmore laboratory (Iverson et al., 2015), contains a series of biological 'parts' for the E. coli chassis; which can be built into transcriptional units using type II restriction cloning. Although the toolkit is a great jumping off point for the rapid assembling of standardised transcriptional units, some useful features, such as C-terminal protein tags, are not included. Before any synthetic BMC constructs could be made, these tags were designed and added to the Toolkit terminators. Another useful feature developed was the modification of the pET28a cloning vector into a Golden-Gate compatible vector to facilitate protein expression-construct generation.

5.2.1 Bacterial microcompartment shell protein constructs

Many efforts have been made to engineer bacterial microcompartment shells, including creating empty carboxysomes (Menon et al., 2008), overexpression of BMC shell proteins from a Pdu microcompartment (Parsons et al., 2008), formation of

nanotubes using a BMC-H protein (Noël et al., 2016) and engineering of BMC shell proteins to introduce a new function (Aussignargues et al., 2016). However, these tend to use classical cloning methods or completely synthetic DNA synthesis. These methods are costly and time consuming, particularly if large numbers of constructs need to be produced.

To quickly generate large numbers of synthetic BMC shells to determine their basic assembly principles, the shell proteins from the *C. phytofermentans* fucose/rhamnose microcompartment and the *Salmonella enterica* Pdu microcompartments were cloned into the MoClo system, to look at how different amounts of the three shell protein domains affect microcompartment formation; and to see what effects, if any, the position of the shell proteins in a 'locus' has on BMC shell formation.

5.2.2 Forward engineering of RBS + aldehyde dehydrogenase localisation sequences to study downstream protein expression

A series of rationally-designed RBS + aldehyde dehydrogenase localisation sequence constructs were made and cloned into the MoClo Toolkit Level 0 plasmid. The RBSs were designed with different binding strengths, at a 1X, 10X, 100X and 1000X proportional log scale. The rationale for these constructs was to identify whether a fixed protein sequence downstream of the designed RBS would a) affect reporter protein expression levels and b) proportionally target the reporter protein to a native microcompartment. With a number of groups attempting this work with related microcompartment proteins, this work is of particular interest in the field (Fan et al., 2010; Kinney et al., 2012; Quin et al., 2016).

After the successful cloning of the RBS+loc constructs into the Toolkit, two different degron tags were added to the 5' region of the MoClo terminators. Several constructs

were made, with either constitutive or controllable promoters, the 1X, 10X, 100X or 1000X RBS+loc, and a GFP or Philov reporter. Constructs were made with and without the C-terminal degron tags.

Although this work was largely preliminary, some initial plate reader and flow cytometry assays suggested that for the Cphy1178 constructs, the 10X RBS+loc gave the highest overall values for GFP fluorescence. The initial hypothesis was that the higher strength RBSs would give proportional GFP expression, however it may be the case that the sequence of the 10X RBS+loc construct is particularly amenable to ribosome binding in *E. coli*. It is important to observe that there are limitations with in silico construct design; bacterial transcription and translation can have many unidentified variables, which would mean that a rational design would not always work correctly in vivo. For this reason, the iterative nature of synthetic biology is useful – as constructs are quick and cheap to clone, many variables can be tested with relatively little effort.

5.2.3 EutM shell proteins – expression and thin-section electron microscopy

Different homologues of the EutM protein show different charge states across the surface of the protein (Held et al., 2013; Pitts et al., 2012), with EutM having a strongly negatively charged pore (Takenoya et al., 2010) and the homologue CD1918 from *Clostridium difficile* having a more even charge across the surface (Pitts et al., 2012). This suggests that there is a difference in charge states across the surface of the protein even between the same protein from different species.

Although TEM images from the four different EutM examples shown do not conclusively show that there is a difference between the protein packing, the fact that

overexpression of the protein forms spirals in itself suggests that the concave and convex sides to the protein have differing charge states, and this could affect the packing of the protein within the spirals. A similar affect has been shown in (Choudhary et al., 2012; Heldt et al., 2009; Parsons et al., 2008) and also with the EutM CD1918 homolgue from *Clostridium phytofermentans* in (Pitts et al., 2012). In terms of BMC production, and the use of shell proteins as scaffolds to increase enzyme efficiency, it means that rational, well-informed design of shell protein structures is needed, with the ability to test large libraries of BMC proteins, which is possible with the MoClo system described here.

5.3 Future work

5.3.1 Propionaldehyde dehydrogenase structure and biochemistry

To further elucidate the role of the E357 residue in the active site of the propionaldehyde dehydrogenase Cphy1178, the NAD^+ turnover assay would be carried out with the E357A variant. Although this particular assay only follows the initial rate-limiting step of NAD^+ to NADH turnover, it would be useful to see if the variant is inactive, along with the H387A and C269A. Soaking Cphy1178 E357A apo-crystals with both NAD^+ and CoA would be beneficial, particularly to try and resolve the pantothenate group that has so far been elusive, or even to capture the acyl- enzyme intermediate.

Another interesting Cphy1178 variant would be to the A235 residue. In non- acylating aldehyde dehydrogenases, this particular residue is a conserved glutamic acid. An A235E variant would possibly show the difference in mechanism between the acylating and non-acylating aldehyde dehdyrogenases.

5.3.2 Bacterial microcompartments as tools for synthetic biology

Overall, bacterial microcompartments confer two main advantages to a bacterial cell; they sequester enzymes and substrates in the case of carboxysomes, increasing enzyme turnover and they protect the cell from toxic and volatile intermediate formed as part of sugar breakdown pathways in metabolosomes. These points could have implications in the production of biofuels (Fan et al., 2010; Lee et al., 2016; Petit et al., 2015; Zuroff and Curtis, 2012) or drug delivery systems (Jahn et al., 2013; Tsai and Yeates, 2011). Interesting directions this project could be taken in are

- constructing and sequencing shell protein transcriptional units on a large scale to look at chimeric and homologous bacterial microcompartment shells, with a view to rationally designing shells for the targeting of specified enzyme pathways
- refining the experimental parameters of the RBS + localisation sequence constructs, to accurately quantify reporter-gene targeting to native microcompartment shells
- to crystallise the EutM homologues described in this study, to get a more accurate view of the different charge states across the protein's surface and how this might be used to design novel protein scaffolding systems

6 Materials and Methods

All chemicals used in this work were purchased from Sigma-Aldrich unless otherwise stated. Plasmid pET-28a(+) was purchased from Novagen. MoClo toolkit was purchased from Addgene (Iverson et al., 2015).

6.1 Aldehyde dehydrogenase enzyme cloning

The full-length wild-type Cphy1178, Cphy1428 and Cphy2642 were first cloned from the *Lachnoclostridium phytofermentans* ISDg (*Clostridium phytofermentans* ISDg). Various constructs including point mutations and truncated constructs were prepared by standard cloning methods as described below.

6.1.1 Full-length and truncated protein cloning

Primers to generate full-length Cphy1178 residues (1-462) were designed between the NcoI and XhoI restriction sites to clone into the pET-28a vector as shown in Table 6.1. The truncated version of Cphy1178 (residues 20-462) was designed to cover the full-length construct, excluding the signal sequence (1-19) and to include restriction enzyme site to clone into the pET-28a vector (Novagen).

To generate each construct, DNA was amplified by PCR in 50 µl reactions typically containing: either 10 ng of genomic template DNA or 1 ng plasmid template DNA; 0.3 µM of both the forward and reverse primer, 0.02 U/µl KOD Hot Start DNA polymerase; 1.5 mM MgSO4; and 0.2 mM (each) of dNTPs in 1 X buffer for KOD Hot Start DNA polymerase. PCR programme was: 95 ºC for 2 min; followed by 30 cycles of 95 ºC for 30 s, annealing temp for 10 s, 70 ºC for 30 s; and a final extension

of 70ºC for 10 min. Annealing temperatures varied between reactions, always being within 5ºC of the melting temperatures of each primer.

The PCR reaction was visualized on a 0.8 % w/v agarose gel and PCR product was purified either by gel extraction or PCR clean-up using the QIAGEN PCR purification or gel extraction kits, following the manufacturer's protocol. Amplified DNA and pET-28a vector were digested with the relevant restriction enzymes following the manufacturers instructions in a typical reaction volume of 20 µl. Digested vector and the digested amplified gene were purified using the QIAGEN PCR purification kit as above and ligated with NEB T4 DNA ligase as per manufacturers instructions. 5 µl ligated plasmid was transformed into 50 µl One Shot TOP10 Chemically Competent *E. coli* (ThermoFisher) with 45 s heat-shock at 42 ºC with recovery in 200 µl LB with incubation at 37 ºC for 40 min. 50 µl of transformed cells were plated onto LB agar supplemented with 50 µg/ml kanamycin and incubated overnight at 37 ºC. Single colonies were used to inoculate 5 ml of LB supplemented with 50 µg/ml kanamycin and were cultured overnight at 37 ºC, shaking at 250 rpm. The plasmid was extracted from overnight cultures using the QIAGEN miniprep kit according to manufacturer's instructions, varying only in DNA elution volume (35 µl). The gene of interest was identified by a Sanger sequencing reaction, carried out by Edinburgh Genomics and using the T7 promoter and T7 terminator primers (Table 6.1).

Construct	Forward primer	Reverse primer	Restriction sites	Tag
Cphy1178$_{(His)}$	**CTCCCATGGGC**ACAGTGAATGAACAATTG	**GGCCTCGAG**TCGGATACACAAACTATC	NcoI/XhoI	C-His
Cphy1178$_{(20-462)}$	**CTCCCATGGGC**CAATTGACACAAACAAAT	**GGCCTCGAG**TTATCGGATACACAAACT	NcoI/XhoI	n/a
Cphy1428$_{(His)}$	**GGCCATATG**GAAAACTTTGATTTTGAT	**GGCCTCGAG**CTACTGACCTTTCATGGC	NdeI/XhoI	N-His
Cphy1428$_{(1-448)}$	**GCGCCATGGGC**TTTGATTTTGATCTGCGT	**GGCCTCGAG**TTATTTCAGGCCAAAGGCAAC	NcoI/XhoI	n/a
Cphy2642$_{(His)}$	**CTCCCATGG**AGTTACAAGAGAAAGAT	**GGCCTCGAG**TATTAGTTGTTTCATAAT	NcoI/XhoI	C-His
Cphy2642$_{(1-452)}$	**GCGCATATG**GAACTGCAAGAAAAAGATCTG	**GCGCTCGAG**TTATTCTTTCACGCCAAAGGC	NdeI/XhoI	N-His
Cphy1178$_{(20-462)}$ C269A	ATAATCTTCCA**GCT**ATTGCAGAAAAAG	CTTTTTCTGCAAT**AGC**TGGAAGATTAT	n/a	n/a
Cphy1178$_{(20-462)}$ H387A	ATGGTAATCGA**GCT**TCCGCACATAT	ATATGTGCGGA**AGC**TCGATTACCAT	n/a	n/a
Cphy1178$_{(20-462)}$ E357A	CCTTTGATTTCTGAAGCGTTAATGATGCCGATTC	GAATCGGCATCATTAACGCTTCAGAAATCAAAGG	n/a	n/a
T7 sequencing primers	**TAATACGACTCACTATAGGG**	**GCTAGTTATTGCTCAGCGG**	n/a	n/a

Table 6.1 Primers used for the aldehyde dehydrogenase cloning

Primer sequences for constructs used in this work. All primers are listed 5' to 3', from left to right. Introduced restriction sites are shown underlined; regions complimentary to genomic DNA shown in bold; sequence mismatches in mutagenic primers are shown in red. All primers are listed 5' to 3', from left to right

6.1.2 Site-directed mutagenesis of Cphy1178$_{(20-462)}$ active site residues

To generate both full-length and truncated mutants Cphy1178(C269A), Cphy1178(H387A) and Cphy1178(E357A), mutagenesis primers were designed using the online programme PrimerX (Table 6.1). The PCR reaction was set up in 50 µl of 1 X PfuUltra buffer and contained: 2.5 pmol/µl each of forward and reverse primer; 40 mM dNTP mix (10 mM each); 2 ng/µl plasmid template (either Cphy1178 or Cphy1178$_{(20-462))}$ and 2 U/µl PfuUltra Hotstart DNA polymerase. The PCR programme was: 95 ºC for 5 min; followed by 18 cycles of 95 ºC for 50 s, 60 ºC for 50 s, 68 ºC for 2.5 min; and a final extension of 68 ºC for 7 min. After PCR, 20 U/µl of the enzyme DpnI was added to the reaction mixture and then incubated at 37 ºC for 60 min. 1 µl of plasmid was transformed in *E. coli*, cultured and the DNA was extracted as described above. The sequence of the mutated gene of interest was identified by a Sanger sequencing reaction, carried out by Edinburgh Genomics using the T7 promoter and T7 terminator primers (Table 6.1)

6.2 Aldehyde dehydrogenase enzyme expression, purification and activity assays

6.2.1 Aldehyde dehydrogenase expression

Single colonies of E. coli BL21(DE3) cells were transformed with protein expression plasmid, and transferred into 10 ml LB medium supplemented with appropriate antibiotic, such as 50 ng/µl kanamycin for pET-28a(+). Inoculated cultures were incubated overnight at 37°C with shaking at 200 rpm. 1 litre of LB medium was spiked with the 10 ml overnight culture and additional antibiotic and incubated at 37°C with shaking at 200 rpm. Recombinant protein expression was induced at OD600 = 0.6 by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and further incubated at 37ºC for four hours. Cells were pelleted by centrifugation at 4000 x g for 30 min at 4°C, and resuspended in 10 times v/w of PBS to wash cells before a second centrifugation at 4000 x g for 15 mins.

6.2.2 Aldehyde dehydrogenase purification

Untagged *C. phytofermentans* BMC shell proteins, and the BMC associated aldehyde dehydrogenase enzymes (Cphy_1178, Cphy_1428 and Cphy_2642), were purified as follows. Cells were resuspended in 10 times v/w of buffer QA (50 mM Tris.HCl pH 8.0) and subjected to lysis by ultrasonication on ice (15 secs on, 15 secs off repeated for 8 cycles) The lysate was clarified by centrifugation (35,000 x g, 30 min) and the supernatant was filtered using a 0.45 µm syringe filter (Millipore). The filtered supernatant was loaded onto a 12 ml Q-sepharose column (GE Healthcare) equilibrated with buffer QA. Unbound sample was washed off with 10 column volumes of buffer QA and protein was eluted with a linear gradient of 0–100% buffer

QB (50 mM Tris.HCl pH 8.0, 1 M NaCl) over 20 column volumes. Peak fractions were analysed by 15% SDS-PAGE and fractions containing the protein of interest were pooled and concentrated using a 10 kDa MWCO centrifugal concentrator (Sartorius). Concentrated samples were subjected to size exclusion chromatography using a Superdex S200 HR16/600 column (GE Healthcare) equilibrated with Size Exclusion Chromatography buffer (50mM Tris.HCl pH8.0, 150mM NaCl). Protein fractions were assessed by 15% SDS-PAGE with molecular weight of fractions estimated with a broad range protein ladder (Fermentas).

6.2.3 Aldehyde dehydrogenase activity assays

Standard aldehyde dehydrogenase assays were performed with the following components: 100 mM Tris.HCl pH 8.0, 100 mM KCl, 0.6 mM β-NAD, 10 mM Beta-mercaptoethanol, and various concentrations of Cphy_1178, Cphy_1428 or Cphy_2642 and various aldehyde substrates. Aldehyde substrates were made up to either 10 mM or 100 mM stock solutions with H2O, to minimise the aldehyde amount needed for each reaction. NAD^+ was the final component added to the reaction, to ensure the shortest time between assay set-up and initial reading. The production of NADH was monitored by measuring change in A340 as a function of time, using an absorption co-efficient of $6.22\,mM^{-1}\,cm^{-1}$ and a pathlength of 0.2 cm. Initial enzymatic activity tests were carried out using a benchtop UV/Vis spectrophotometer V-730 UV/VIS spectrophotometer (JASCO Inc.) (pathlength 1 cm), with 1 ml reaction volume, to test whether purified proteins were active.

For purified proteins with activity from initial exploratory experiments, assays were carried out to determine their kinetic parameters using a 96-well micro-titre plate reader (SpectraMax M5 (Molecular Devices)) with each well containing 300 µl of reaction buffer. For the acetaldehyde assay, Cphy_1178 was used at 100 nM final concentration, with A340 measured every 7 s for 182 s and the assay was carried out in 5 replicates. The assay blank used standard assay components without protein. Initial rates from linear regression for NADH produced for different substrate concentrations were obtained from data analysis. These initial rates produced a curve fitted with non-linear regression applying the Michaelis-Menten equation using Prism 6 software (Graphpad). The nonlinear regression model associated with the Michaelis-Menten equation is as follows:

$$V = \mathrm{C(Vmax)/\ (C + Km)} + \varepsilon$$

where ε represents normally distributed errors, with zero mean and constant variance $\sigma 2$. It provides confidence intervals, estimates, and statistical hypothesis tests based on this assumption.

6.3 Crystallisation/Optimisation

6.3.1 Protein crystallisation screening and optimisation

Proteins were screened for crystallisation against the following commercial crystallisation kits: JCSG+, Structure Screen, and MIDAS from Molecular Dimensions. An Art Robbins Gryphon nanolitre pipetting robot was used to dispense 65 µl of each crystallisation screen from a master screen kit into a MRC 96-well crystallisation plate. 0.1 µl reservoir solution was pipetted into each well, followed by 0.2 µl and 0.1 µl protein dispensed into the top and bottom wells by the nanodispenser. The sitting- drop vapour-diffusion crystallisation plate was sealed with the X-Seal Manual Adhesive Sealer XCS-384 (FluidX).

Purified Cphy1178(20-462) and the C269A, H387A and E357A variants were concentrated to between 8-12 mg/ml in 30 mM NaCl, 50 mM Tris.HCl pH8.0 and crystallised by sitting drop vapour diffusion in drops of 2 µl protein plus 2 µl crystallization solution, over 1 ml of crystallization solution after finding appropriate conditions from JCSG+ screen. Crystal trays were incubated at 18 ºC for between one and three weeks, and crystal formation was monitored using a Leica DFC450 C digital microscope at 10 X resolution.

6.3.2 Crystal soaking and mounting

Crystals were obtained in 0.1 M sodium acetate pH4.7, 1.8 M ammonium sulfate and grew as bi-pyramids of 50-100 µm in size. Crystals were harvested from the well using a LithLoop (Molecular Dimensions), transferred briefly to a saturated ammonium sulfate solution and then paratone oil (Molecular Dimensions). Excess oil was wicked off and the cryo-protected crystals were flash-cooled in liquid nitrogen. Crystals of

Cphy1178(20-462) with NAD+ and CoA ligands were obtained by soaking harvested crystals in a solution of 0.1 M sodium acetate pH4.7, 1.8 M ammonium sulfate, supplemented with 10 mM ligand and cryo-protected as described above.

6.3.3 X-ray diffraction data collection and processing

All crystallographic datasets were collected on beamlines I02 and I04 at Diamond Light Source (Didcot, UK) at 100 K and using ADSC CCD, or Pilatus 6 M detectors. Diffraction data were integrated and scaled using XDS (Kabsch, 2010) and merged with Aimless (Evans and Murshudov, 2013). The crystallographic point group and space group were determined by Pointless and symmetry related reflections were merged with Aimless (Evans and Murshudov, 2013). The resolution cut-off used for structure determination and refinement was determined based on the completeness (>95%), I/sigI (>1.5) and CC1/2 (>0.5) criterion of Karplus and Diederichs (Karplus and Diederichs, 2012).

6.3.4 Crystallographic model building and refinement

The structure of Cphy1178_20 was determined by molecular replacement using the PDD homologue (PDB ID: 3K9D) as the search model. A single solution was found by molecular replacement using Phaser MR (McCoy et al., 2007). The initial model was rebuilt using Phenix.autobuild (Terwilliger et al., 2008) followed by cycles of refinement with Phenix.refine (Adams et al., 2010), with manual rebuilding and model inspection in Coot (Emsley et al., 2010). Each refinement cycle in Phenix.refine included ten macrocycles of bulk solvent correction, anisotropic scaling, coordinate refinement, individual B-factors and occupancy refinement. The refinement target

weights included scale factor for X-ray/stereochemistry weight (wxc_scale) as 0.5; for Xray/ADP weight (wxu_scale) as 1.0; stereochemistry weight scale (wc) as 1.0; ADP weight scale (wu) as 1.0. The refinement program was repeated for several cycles until the Rwork, Rfree and stereo chemical restraints achieved convergence. In the final stage of the refinement, water molecules and hydrogens were added to the protein model automatically in Phenix and further inspected in Coot manually. The final model was also refined with torsional NCS restraints (2.5° for restraint sigma and 15° for restraint limit). The crystal structures of Cphy1178$_{(20\text{-}462)}$ C269A, H387A and E357A produced as above, however using PDBID:4C3S as the molecular replacement model.

6.4 Cidar MoClo cloning

6.4.1 Insertion of shell protein constructs into Level 0 vector

The CIDAR MoClo Parts Kit was a gift from Douglas Densmore (Addgene kit # 1000000059). All Level 0 plasmids have an ampicillin resistance gene, and cloning into any Level 0 plasmid was carried out with the BpiI type II restriction enzyme, and T4 DNA ligase, with quantities of each specified below.

Shell protein constructs (Table 6.2) were designed with a 5' C (AATG) overhang and a 3' D (ACCT) overhang to be inserted into the MoClo toolkit system. The Cphy_1182 construct had an internal BpiI site, so this was mutated via site-directed mutagenesis as in section 6.1.2 prior to Level 0 cloning. Genes were amplified by PCR from plasmid containing the gene of interest (supplied by Dr Jon Marles-Wright) and purified as in section 6.1.1, followed by a one-tube digestion/ligation reaction.

Construct	Forward primer	Reverse primer	5’ Fusion site	3’ Fusion site
Cphy_1176	GGCGAAGACATAatgacaaccgaagataaacttcgcattg	GGCGAAGACATACCTggtacgtgttatgtcacat	AATG	ACCT
Cphy_1180	GGCGAAGACATAatggtttagcagttggatttta	GGCGAAGACATACCTattcttatcaaagccgcttaagtg	AATG	ACCT
Cphy_1181	GGCGAAGACATAatggaaaatcgttaggctttatt	GGCGAAGACATACCTcatttcttcgttttgtggctt	AATG	ACCT
Cphy_1182	GGCGAAGACATAatggatgataaattagat	GCCGAAGACATACCTtttaattcctggtaaaat	AATG	ACCT
Cphy_1184	GGCGAAGACATAatgttaatcggcaaagtaa	GGCGAAGACATACCTctccagtcctgttgcatc	AATG	ACCT
Cphy_1186	GGCGAAGACATAatgagtaaagcaattgga	GGCGAAGACATACCTcatcaacttattgcgtaa	AATG	ACCT
STM2038	GGCGAAGACATAatgcaacaagaagcactagga	GGCGAAGACATACCTttggctaattcccttcgg	AATG	ACCT
STM2039	GGCGAAGACATAatgagcagcaatgagctg	GGCGAAGACATACCTgatgtaggacggacgatc	AATG	ACCT
STM2045	GGCGAAGACATAatgaataacgcactggga	GGCGAAGACATACCTggctgatttcggtaaaat	AATG	ACCT
STM2046	GGCGAAGACATAatgaagcaatcactggga	GGCGAAGACATACCTcgcttcacctcgcttgcc	AATG	ACCT
STM2049	GGCGAAGACATAatgcatctggcacgagtc	GGCGAAGACATACCTacacgaaagcgtatctac	AATG	ACCT
STM2054	GGCGAAGACATAatgtctcaggctatagga	GGCGAAGACATACCTcccctccaccatctgtcg	AATG	ACCT
STM2055	GGCGAAGACATAatggaaagacaaccgaca	GGCGAAGACATACCTcgtccgggtgatcgagca	AATG	ACCT
Cphy1182_mut	GTGCAGCTAGCGCTGGGAGACTTGGTGAGTTAG	CTAACTCACCAAGTCTCCCAGCGCTAGCTGCAC	n/a	n/a

Table 6.2 Primers used for shell protein to Level 0 cloning and

Primer sequences for constructs used in this work. All primers are listed 5’ to 3’, with overhangs highlighted in red. All primers are listed 5’ to 3’, from left to right. Cphy_1182 mutated region is highlighted in blue

The reaction components and digestion/ligation parameters are shown in Tables 6.3 and 6.4 respectively. 5 µl ligated plasmid was transformed into 50 µl One Shot TOP10 Chemically Competent *E. coli* (ThermoFisher) with 45 s heat-shock at 42 ºC with recovery in 200 µl LB with incubation at 37 °C for 40 min

Component	Amount
Shell protein part	10 fmol
NEB Ligase buffer	1 X
T4 DNA ligase (NEB)	20 U/reaction
BsaI (NEB)	10 U/reaction
Total reaction volume 20 µl	

Table 6.3 Level 0 Golden-Gate cloning reaction components and volumes

Step number	Temperature (ºC)	Time (mins)
1	37	20
2	37	1.5
3	16	3
Cycle Step 2 and 3 x 7		
4	50	5
5	80	10

Table 6.4 Level 0 Golden-Gate cloning digestion and ligation parameters

Transformed cells were plated on to X-gal plates (2.5 ml of 20 mg/ml stock solution of X-gal (200 mg X-gal, 10 ml DMSO), 1 mM IPTG, 50 µg/ml ampicillin) and

screened for white colonies. Single colonies were incubated in 10 ml LB supplemented with appropriate antibiotic for plasmid propagation. Plasmids were purified from the cells using the Qiagen Miniprep kit based on the manufacturer instructions. Gene insertion into the Level 0 plasmid was confirmed by Sanger sequencing of the resulting plasmid (Edinburgh Genomics).

6.4.2 Insertion of terminator tags into MoClo toolkit

The following tags were cloned into empty Level 0 DE, DF, DG and DH plasmids: Flag tag, HA tag, Strep tag and His6 tag. PCR primer sequences are shown in Table 6.5.

Tag name	Primer sequence
B0015_FlagSTOP_fwd	GGCGAAGACATAGGTCTgattacaaggatgacgatgacaagtaaccaggcatcaaataaaacgaa
B0015_HASTOP_fwd	GGCGAAGACATAGGTCTtacccatacgacgtcccagactacgcttaaccaggcatcaaataaaacgaa
B0015_StrepSTOP_fwd	GGCGAAGACATAGGTCTtggagccacccccagttcgagaagtaaccaggcatcaaataaaacgaa
B0015_HisSTOP_fwd	GGCGAAGACATAGGTCTctagagcatcatcaccatcaccattaaccaggcatcaaataaaacgaa
B0015_STOP_fwd	GGCGAAGACATAGGTAAccaggcatcaaataaaacgaa
B0015_DH_rev	GGCGAAGACATTAGTtataaacgcagaaaggcc
B0015_DG_rev	GGCGAAGACATGGCAtataaacgcagaaaggcc
B0015_DF_rev	GGCGAAGACATAGCGtataaacgcagaaaggcc
B0015_DE_rev	GGCGAAGACATAAGCtataaacgcagaaaggcc

Table 6.5 Forward and reverse primers for terminator tag cloning
B0015_ refers to the terminator gene name in the MoClo toolkit, DE, DF, DG and DH are the MoClo specified overhangs.

PCR products were cloned using the MoClo Toolkit terminator, B0015, as a plasmid template (described in section 6.1.2). PCR products were subsequently purified and cloned into the appropriate Level 0 vector as described above.

6.5 RBS + localisation sequence design and cloning

6.5.1 RBS + localisation sequence design

Forward engineered RBSs were designed using the RBS calculator software from the Salis lab (Espah Borujeni et al., 2014; Salis et al., 2009), using the residues 1-18 in the STM2051 construct and 1-19 in the Cphy1178 construct. The pre-sequence was set as the MoClo B' overhang (TACT), and the protein coding sequence was set as the first 18 and 19 aa residues of STM2051 and Cphy1178 respectively. Proportional scales were set at 1X, 10X, 100X and 1000X strength for both constructs.

6.5.2 RBS+localisation sequence cloning

Forward and reverse primers for each construct were re-suspended in nuclease-free water to a concentration of 100 μM, and added together in NEB fast-digest buffer in equimolar amounts. The primer mix was placed in a heat block at 95 ºC for 5 minutes, with subsequent cooling at room temperature. Resulting annealed primers were then used as the component parts and cloned into the Level 0 vector, as shown in section 6.1.2.

Multiple Level 1 transcriptional units were generated using a mix of toolkit parts and lab-designed RBS+loc constructs and engineered terminators as shown in Table 6.7. Sheet showing each toolkit part relative to the kit number is shown in the Appendix.

Construct	Forward primer	Reverse primer
S 1X	…AAAATAAGAACAGAACCATTCAGCCCCAGGGatgaatacttctg aactcgaaaccctgattcgcaccattcttagcgagcaatta	…taattgctcgctaagaatggtgcgaatcagggtttcgagttcagaagtattca tCCCTGGGGCTGAATGGTTCTGTTCTTATTTT
S 10X	…GAACTCGCGCCGCTCAACGCGAGTTCAAGTCAGatgaatactt ctgaactcgaaaccctgattcgcaccattcttagcgagcaatta	…taattgctcgctaagaatggtgcgaatcagggtttcgagttcagaagtattcat CTGACTTGAACTCGCGTTGAGCGGCGCGAGTTC
S 100X	…AGGTCAGCGCGTAATTAAACCGCTTCCGCTACAGatgaatact tctgaactcgaaaccctgattcgcaccattcttagcgagcaatta	…taattgctcgctaagaatggtgcgaatcagggtttcgagttcagaagtattcat CTGTAGCGGAAGCGGTTTAATTACGCGCTGA CCT
S 1000X	…CGCTATTTCACTATCCAAATACCGAGACACTTatgaatacttctg aactcgaaaccctgattcgcaccattcttagcgagcaatta	…taattgctcgctaagaatggtgcgaatcagggtttcgagttcagaagtattcat AAGTGTCTCGGTATTTGGATAGTGAAATAG CG
C 1X	…TTCCCCAAACACTTCATTAGTCACAGCAAATatgacagtgaatg aacaattggttcaagacatcattaaaaatgtcgtagcgagtatg	…catactcgctacgacatttttaatgatgtcttgaaccaattgttcattcactgtcat ATTTGCTGTGACTAATGAAGTGTTTGGGG AA
C 10X	…GGAGCGGGCCGTCATAAAAGAAAAAATCTACTatgacagtga atgaacaattggttcaagacatcattaaaaatgtcgtagcgagtatg	…catactcgctacgacatttttaatgatgtcttgaaccaattgttcattcactgtcat AGTAGATTTTTTCTTTTATGACGGCCCGCT CC
C 100X	…TCTACGGCCTCCAGTCAGATAAAGTAGACGTAatgacagtgaa tgaacaattggttcaagacatcattaaaaatgtcgtagcgagtatg	…catactcgctacgacatttttaatgatgtcttgaaccaattgttcattcactgtcat TACGTCTACTTTATCTGACTGGAGGCCGTA GA
C 1000X	…CTAGTACTATAAACCTACAAAGACATCAGAATATatgacagt gaatgaacaattggttcaagacatcattaaaaatgtcgtagcgagtatg	…catactcgctacgacatttttaatgatgtcttgaaccaattgttcattcactgtcat ATATTCTGATGTCTTTGTAGGTTTATAGTAC TAG

Table 6.6 Forward and reverse PCR primers for RBS + localisation sequence constructs
The protein coding sequence is shown in lowercase letters and the 5' and 3' overhangs with junk DNA are shown in uppercase letters. The 3 red dots refer to sequence GGCGAAGACATTACT and the three blue dots refer to sequence GGCGAAGACATCATT. S is STM2051 and C is Cphy1178

Component	Amount
Shell protein part	10 fmol
NEB Ligase buffer	1 X
T4 DNA ligase (NEB)	20 U/reaction
BsaI (NEB)	10 U/reaction
Total reaction volume 20 µl	

Table 6.7 Level 1 Golden-Gate cloning reaction components and volumes

Step number	Temperature (ºC)	Time (mins)
1	37	1.5
2	16	3
Cycle Step 1 and 2 x 15		
3	50	5
3	80	10

Table 6.8 Level 1 Golden-Gate cloning digestion and ligation parameters

TU	Promoter_AB	RBS+loc_BC	Reporter_CD	Terminator	Level 1 plasmid_AE
1	R0010	All	E0040m	STOP	DVK_AE
2	R0010	All	E0040m	STOP	DVK_AE
3	J23100	All	E0040m	STOP	DVK_AE
4	J23100	All	E0040m	STOP	DVK_AE
5	J23100	All	E0040m	LAA_Degron tag	DVK_AE
6	J23100	All	E0040m	LVA_Degron tag	DVK_AE
7	J23100	All	Philov	STOP	DVK_AE
8	J23100	All	E0040m	LVA_Degron tag	DVK_AE
9	J23100	All	E0040m	LVA_Degron tag	DVK_AE

Table 6.9 Level 1 RBS+loc transcriptional unit constructs

6.5.3 RBS + localisation sequence plate reader assays

Fluorescence levels of gene expression were assayed by fluorometry at cell population level and by flow cytometry at single-cell level. Cells grown in 96-well plates were monitored and assayed using a BMG POLARstar fluorometer for repeated absorbance (OD600) and fluorescence (485 nm for excitation, 520 ± 10 nm for emission, Gain = 1,000) readings (20 min per cycle).

6.5.4 RBS+ localisation sequence flow cytometry

For each variant, cell cultures were diluted 1:4000 in 200 μL of phosphate buffered saline in 96-well plates, and 5000 events were collected on a Millipore Guava easyCyte 5HT flow cytometer. Gates were set around the cell population using the forward and side scatter channels, and average fluorescence values were calculated using the geometric mean.

6.6 EutM cloning, protein expression and purification and thin-section electron microscopy

6.6.1 Cloning EutM into Golden-Gate modified pET28a

Cloning was carried out as shown in earlier sections, replacing the Level 0 CD plasmid with the pET28a-GG-RFP plasmid Level 1 plasmid. BsaI and T4 DNA ligase were used consistently for the Golden-Gate cloning reactions, the quantities shown in section 6.1.4 are the same throughout the project.

EutM protein *spp*	Forward primer	Reverse primer	Reverse $His_{6\ tag}$
D. psychrophila	GGCGGTCTCAAATGGATTCATT AGGAATGATTGAA	GGCGGTCTCAACCTCTATTTGTCACGT TTGAAGGG	GGCGGTCTCAACCTTTGTCACGTTTGAA GGGAAA
P. hadalis	GGCGGTCTCAAATGGACGCTTT AGGTATTTTA	GGCGGTCTCAACCTTTACTTTTTTGCT GTGATAG G	GGCGGTCTCAACCTCTTTTTTGCTGTGAT AGGAA A
S. enterica	GGCGGTCTCAAATGGAAGCATT AGGAATGA	GGCGGTCTCAACCTTCAAATGTTGCT GTCGCCTT T	GGCGGTCTCAACCTAATGTTGCTGTC GCCTTTGA A
T. linaloolentis	GGCGGTCTCAAATGGAAGCCCT GGGACTGATC	GGCGGTCTCAACCTTCAGTCCAGTCCG CTTTCCA T	GGCGGTCTCAACCTGTCCAGTCCGCTTTC CATCT T

Table 6.11 Primers for EutM constructs to be cloned into the amended Golden-Gate pET28a vector
C' overhangs are highlighted in red and D' overhangs are highlighted in blue

6.6.2 Ammonium sulfate precipitation of *D. psychrophila* and *P. hadalis* EutM proteins

Dp and Ph proteins were purified by anion exchange, as described in section 6.1.2. The supernatant was centrifuged at 5000 x g for 15 minutes at room temperature. Proteins were sequentially precipitated from 50 ml of crude extract by stepwise addition of solid ammonium sulfate with stirring at a certain degree of saturation, and then centrifuged again at 5000 x g at 4°C for 15 minutes. The pellet obtained after each centrifugation was resuspended in max. 10 ml anion exchange buffer. Steps above were repeated for 20, 30, 50 and 80% ammonium sulfate.

Precipitated fractions were analyzed for its protein purity by 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

6.6.3 Thin-section electron microscopy of EutM constructs

E. coli Top10 cells transformed with the plasmids for untagged and His6-tagged Dp, Ph, Stm and Tl constructs were grown to mid-log phase in Luria-Bertani media supplemented with 50 µg/ml kanamycin. 1 ml of these cells were fixed in 2.5% (v/v) glutaraldehyde, 50 mM sodium cacodylate pH 7.0 for 30 mins at room temperature. Cells were subsequently immobilized in 2% (w/v) water-agar and post-fixed in 1.5% (w/v) osmium tetroxide in 50 mM sodium cacodylate pH 7.0 for 1 hour at 4°C followed by dehydration in an ethanol series. The final 70% dehydration step was supplemented with 1% (w/v) uranyl acetate and was performed overnight at 21°C. 100 nm sections were post-stained with 2% (w/v) uranyl acetate and analysed using a Philips CM100 transmission electron microscope.

7 Appendix

7.1 Calibration of HiLoad 16/600 Superdex 200pg Size Exclusion Chromatography column

Size Exclusion Chromatography provides a means to estimate the molecular weight or size (Stokes radius) of proteins under certain buffer and temperature conditions. The molecular weight of unknown proteins can be determined by comparing the elution volume related parameter Kav to those calculate from known calibration standards run on the same column. Here a calibration curve for a HiLoad 16/600 Superdex 200pg column was prepared by running LMW/HMW calibration kits (GE Healthcare) to measure the elution volume (Ve) of each protein standard (Figure 1 A). The void volume (Vo) of S200 column was estimated based on the retention volume of blue dextran 2000. The Kav value was derived based on the equation:

$$Kav = (Ve – Vo) / (Vt – Vo)$$

Where Vt equals to the total bed volume of particular column. Kav was then plotted against the logarithm of molecular weight of protein standards (Figure 1 B) It should be borne in mind that protein shape also affects the Kav- based molecular weight estimation and the assumption that the protein samples are approximately spherical.

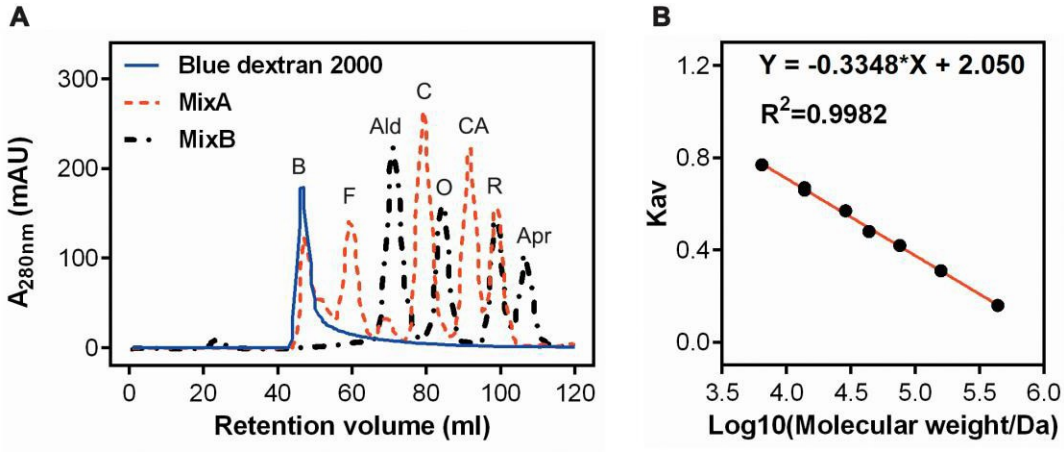


Figure 1 Chromatographic separation and calibration curve for HiLoad 16/600 Superdex 200 pg column.
(A) Chromatogram of HiLoad 16/600 Superdex 200 pg Size Exclusion Chromatography column calibration using blue dextran 2000 (B, blue solid curve), MixA (red dashed curve) and MixB (black dashed curve) in which blue dextran determines the void volume. MixA contains ferritin (F, 440 kDa), conalbumin (C, 75 kDa), carbonic anhydrase (CA, 29 kDa) and ribonuclease A (R, 13.7 kDa). MixB consists of ovalbumin (O, 44 kDa), aprotinin (Apr, 6.5 kDa), ribonuclease A and aldolase (Ald, 158 kDa). (B) Calibration curve of HiLoad 16/600 Superdex 200 pg column.

8 Bibliography

Adams, P.D., Afonine, P. V, Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D. Biol. Crystallogr. *66*, 213–221.

Aussignargues, C., Pandelia, M.-E., Sutter, M., Plegaria, J.S., Zarzycki, J., Turmo, A., Huang, J., Ducat, D.C., Hegg, E.L., Gibney, B.R., et al. (2016). Structure and Function of a Bacterial Microcompartment Shell Protein Engineered to Bind a [4Fe-4S] Cluster. J. Am. Chem. Soc. *138*, 5262–5270.

Axen, S.D., Erbilgin, O., and Kerfeld, C.A. (2014). A Taxonomy of Bacterial Microcompartment Loci Constructed by a Novel Scoring Method. PLoS Comput. Biol. *10*, e1003898.

Badger, M.R., and Price, G.D. (1992). The CO_2 concentrating mechanism in cyanobactiria and microalgae. Physiol. Plant. *84*, 606–615.

Bateman, O.A., Purkiss, A.G., Van Montfort, R., Slingsby, C., Graham, C., and Wistow, G. (2003). Crystal structure of η-crystallin: Adaptation of a class 1 aldehyde dehydrogenase for a new role in the eye lens. Biochemistry *42*, 4349–4356.

Becher, E., Heese, A., Claußen, L., Eisen, S., Jehmlich, N., Rohwerder, T., and Purswani, J. (2018). Active site alanine preceding catalytic cysteine determines unique substrate specificity in bacterial CoA-acylating prenal dehydrogenase. FEBS Lett.

Bobik, T.A., Havemann, G.D., Busch, R.J., Williams, D.S., and Aldrich, H.C. (1999). The propanediol utilization (pdu) operon of *Salmonella enterica serovar Typhimurium* LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B(12)-dependent 1, 2-propanediol degradation. J. Bacteriol. *181*, 5967–5975.

Bonacci, W., Teng, P.K., Afonso, B., Niederholtmeyer, H., Grob, P., Silver, P.A., and Savage, D.F. (2012). Modularity of a carbon-fixing protein organelle. Proc. Natl. Acad. Sci. U. S. A. *109*, 478–483.

Brinsmade, S.R., Paldon, T., and Escalante-Semerena, J.C. (2005). Minimal functions and physiological conditions required for growth of *Salmonella enterica* on ethanolamine in the absence of the metabolosome. J. Bacteriol. *187*, 8039–8046.

Cai, F., Menon, B.B., Cannon, G.C., Curry, K.J., Shively, J.M., and Heinhorst, S. (2009). The Pentameric Vertex Proteins Are Necessary for the Icosahedral Carboxysome Shell to Function as a CO_2 Leakage Barrier. PLoS One *4*, e7521.

Cai, F., Sutter, M., Bernstein, S.L., Kinney, J.N., and Kerfeld, C.A. (2015). Engineering Bacterial Microcompartment Shells: Chimeric Shell Proteins and Chimeric Carboxysome Shells. ACS Synth. Biol. *4*, 444–453.

Cai, F., Bernstein, S.L., Wilson, S.C., and Kerfeld, C.A. (2016). Production and Characterization of Synthetic Carboxysome Shells with Incorporated Luminal Proteins. Plant Physiol. *170*, 1868-77

Cameron, J.C., Wilson, S.C., Bernstein, S.L., and Kerfeld, C.A. (2013). Biogenesis of a bacterial organelle: the carboxysome assembly pathway. Cell *155*, 1131–1140.

Casini, A., MacDonald, J.T., De Jonghe, J., Christodoulou, G., Freemont, P.S., Baldwin, G.S., and Ellis, T. (2014). One-pot DNA construction for synthetic biology: the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy. Nucleic Acids Res. *42*, e7.

Chen, A.H., and Silver, P.A. (2012). Designing biological compartmentalization. Trends Cell Biol. *22*, 662–670.

Cheng, S., Liu, Y., Crowley, C.S., Yeates, T.O., and Bobik, T.A. (2008). Bacterial microcompartments: their properties and paradoxes. Bioessays *30*, 1084–1095.

Choudhary, S., Quin, M.B., Sanders, M.A., Johnson, E.T., and Schmidt-Dannert, C. (2012). Engineered protein nano-compartments for targeted enzyme localization. PLoS One *7*, e33342.

Chowdhury, C., Sinha, S., Chun, S., Yeates, T.O., and Bobik, T.A. (2014). Diverse bacterial microcompartment organelles. Microbiol. Mol. Biol. Rev. *78*, 438–468.

Cleland, W.. (1963). The kinetics of enzyme-catalyzed reactions with two or more substrates or products: I. Nomenclature and rate equations. Biochim. Biophys. Acta - Spec. Sect. Enzymol. Subj. *67*, 104–137.

Crowley, C.S., Sawaya, M.R., Bobik, T.A., and Yeates, T.O. (2008). Structure of the PduU Shell Protein from the Pdu Microcompartment of *Salmonella*. Structure *16*, 1324–1332.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. U. S. A. *97*, 6640–6645.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D. Biol. Crystallogr. *66*, 486–501.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One *3*, e3647.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS One *4*, e5553.

Engler, C., Youles, M., Gruetzner, R., Ehnert, T.-M., Werner, S., Jones, J.D.G., Patron, N.J., and Marillonnet, S. (2014). A golden gate modular cloning toolbox for plants. ACS Synth. Biol. *3*, 839–843.

Erbilgin, O., Sutter, M., and Kerfeld, C.A. (2016). The Structural Basis of Coenzyme A Recycling in a Bacterial Organelle. PLOS Biol. *14*, e1002399.

Espah Borujeni, A., Channarasappa, A.S., and Salis, H.M. (2014). Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. Nucleic Acids Res. *42*, 2646–2659.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr. D. Biol. Crystallogr. *69*, 1204–1214.

Fan, C., Cheng, S., Liu, Y., Escobar, C.M., Crowley, C.S., Jefferson, R.E., Yeates, T.O., and Bobik, T.A. (2010). Short N-terminal sequences package proteins into bacterial microcompartments. Proc. Natl. Acad. Sci. U. S. A. *107*, 7509–7514.

Forouhar, F., Kuzin, A., Seetharaman, J., Lee, I., Zhou, W., Abashidze, M., Chen, Y., Yong, W., Janjua, H., Fang, Y., et al. (2007). Functional insights from structural genomics (Springer Netherlands). *8*, 37-44

Gitai, Z. (2005). The New Bacterial Cell Biology: Moving Parts and Subcellular Architecture. Cell *120*, 577–586.

Gouet, P., Robert, X., and Courcelle, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res. *31*, 3320–3323.

Held, M., Quin, M.B., and Schmidt-Dannert, C. (2013). Eut bacterial microcompartments: insights into their function, structure, and bioengineering applications. J. Mol. Microbiol. Biotechnol. *23*, 308–320.

Heldt, D., Frank, S., Seyedarabi, A., Ladikis, D., Parsons, J.B., Warren, M.J., and Pickersgill, R.W. (2009). Structure of a trimeric bacterial microcompartment shell protein, EtuB, associated with ethanol utilization in *Clostridium kluyveri*. Biochem. J. *423*, 199-207

Huseby, D.L., and Roth, J.R. (2013). Evidence that a metabolic microcompartment contains and recycles private cofactor pools. J. Bacteriol. *195*, (12) 2864-2879

Iancu, C. V., Morris, D.M., Dou, Z., Heinhorst, S., Cannon, G.C., and Jensen, G.J. (2010). Organization, Structure, and Assembly of α-Carboxysomes Determined by Electron Cryotomography of Intact Cells. J. Mol. Biol. *396*, 105–117.

Iverson, S., Haddock, T.L., Beal, J., and Densmore, D. (2015). CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enables Rapid Combinatorial Design for Synthetic and Traditional Biology. ACS Synth. Biol.*5*, (1) 99-103

Jahn, D., Wittmann, C., Frank, S., Lawrence, A.D., Prentice, M.B., and Warren, M.J. (2013). Bacterial microcompartments moving into a synthetic biological world. J. Biotechnol. *163*, 273–279.

Jakobson, C.M., Tullman-Ercek, D., Slininger, M.F., and Mangan, N.M. (2017). A systems-level model reveals that 1,2-Propanediol utilization microcompartments enhance pathway flux through intermediate sequestration. PLOS Comput. Biol. *13*, e1005525.

Jeter, R.M. (1990). Cobalamin-dependent 1,2-propanediol utilization by *Salmonella typhimurium*. J. Gen. Microbiol. *136*, 887–896.

Jorda, J., Lopez, D., Wheatley, N.M., and Yeates, T.O. (2013). Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. Protein Sci. *22*, 179–195.

Kabsch, W. (2010). XDS. Acta Crystallogr. D. Biol. Crystallogr. *66*, 125–132.

Karplus, P.A., and Diederichs, K. (2012). Linking crystallographic model and data quality. Science *336*, 1030–1033.

Kerfeld, C.A., and Melnicki, M.R. (2016). Assembly, function and evolution of cyanobacterial carboxysomes. Curr. Opin. Plant Biol. *31*, 66–75.

Kerfeld, C.A., Sawaya, M.R., Tanaka, S., Nguyen, C. V., Phillips, M., Beeby, M., and Yeates, T.O. (2005). Protein Structures Forming the Shell of Primitive Bacterial Organelles. Science (80-.). *309*.

Kerfeld, C.A., Heinhorst, S., and Cannon, G.C. (2010). Bacterial Microcompartments. Annu. Rev. Microbiol. *64*, 391–408.

Kerfeld, C.A., Aussignargues, C., Zarzycki, J., Cai, F., and Sutter, M. (2018). Bacterial microcompartments. Nat. Rev. Microbiol.

Kinney, J.N., Salmeen, A., Cai, F., and Kerfeld, C.A. (2012). Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. J. Biol. Chem. *287*, 17729–17736.

Kofoid, E., Rappleye, C., Stojiljkovic, I., and Roth, J. (1999). The 17-gene ethanolamine (eut) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins. J. Bacteriol. *181*, 5317–5329.

Lassila, J.K., Bernstein, S.L., Kinney, J.N., Axen, S.D., and Kerfeld, C.A. (2014). Assembly of Robust Bacterial Microcompartment Shells using Building Blocks from an Organelle of Unknown Function. J. Mol. Biol. *426,* (11) 2217-28

Lawrence, A.D., Frank, S., Newnham, S., Lee, M.J., Brown, I.R., Xue, W.-F., Rowe, M.L., Mulvihill, D.P., Prentice, M.B., Howard, M.J., et al. (2014). Solution Structure of a Bacterial Microcompartment Targeting Peptide and Its Application in the Construction of an Ethanol Bioreactor. ACS Synth. Biol. *3,*(7):454-465

Leal, N. a., Havemann, G.D., and Bobik, T. a. (2003). PduP is a coenzyme-a-acylating propionaldehyde dehydrogenase associated with the polyhedral bodies involved in B12-dependent 1,2-propanediol degradation by *Salmonella enterica serovar Typhimurium* LT2. Arch. Microbiol. *180*, 353–361.

Lee, M.J., Brown, I.R., Juodeikis, R., Frank, S., and Warren, M.J. (2016). Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in Escherichia coli. Metab. Eng. *36*, 48-56

Lei, Y., Pawelek, P.D., and Powlowski, J. (2008). A Shared Binding Site for NAD $^{+}$ and Coenzyme A in an Acetaldehyde Dehydrogenase Involved in Bacterial Degradation of Aromatic Compounds †. Biochemistry *47*, 6870–6882.

Liu, Z.-J., Sun, Y.-J., Rose, J., Chung, Y.-J., Hsiao, C.-D., Chang, W.-R., Kuo, I., Perozich, J., Lindahl, R., Hempel, J., et al. (1997). The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. Nat. Struct. Biol. *4*, 317–326.

Machado, I.M.P., and Atsumi, S. (2012). Cyanobacterial biofuel production. J. Biotechnol. *162*, 50–56.

Marsischky, G., and LaBaer, J. (2004). Many paths to many clones: a comparative look at high-throughput cloning methods. Genome Res. *14*, 2020–2028.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

Menon, B.B., Dou, Z., Heinhorst, S., Shively, J.M., and Cannon, G.C. (2008). *Halothiobacillus neapolitanus* carboxysomes sequester heterologous and chimeric RubisCO species. PLoS One *3*, e3570.

Methé, B.A., Leschine, S.B., and Warnick, T.A. (2002). *Clostridium phytofermentans* sp. nov., a cellulolytic mesophile from forest soil. Int. J. Syst. Evol. Microbiol. *52*, 1155–1160.

Moore, S.J., Lai, H.-E., Kelwick, R.J.R., Chee, S.M., Bell, D.J., Polizzi, K.M., and Freemont, P.S. (2016). EcoFlex: A Multifunctional MoClo Kit for *E. coli* Synthetic Biology. ACS Synth. Biol. *5*, 1059–1069.

Noël, C.R., Cai, F., and Kerfeld, C.A. (2016). Purification and Characterization of Protein Nanotubes Assembled from a Single Bacterial Microcompartment Shell Subunit. Adv. Mater. Interfaces *3*, 1500295.

Osbourn, A.E., O'Maille, P.E., Rosser, S.J., and Lindsey, K. (2012). Synthetic biology. New Phytol. *196*, 671–677.

Parsons, J.B., Dinesh, S.D., Deery, E., Leech, H.K., Brindley, A.A., Heldt, D., Frank, S., Smales, C.M., Lünsdorf, H., Rambach, A., et al. (2008). Biochemical and structural insights into bacterial organelle form and biogenesis. J. Biol. Chem. *283*, 14366–14375.

Parsons, J.B., Frank, S., Bhella, D., Liang, M., Prentice, M.B., Mulvihill, D.P., and Warren, M.J. (2010). Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. Mol. Cell *38*, 305–315.

Penrod, J.T., and Roth, J.R. (2006). Conserving a volatile metabolite: a role for carboxysome-like organelles in Salmonella enterica. J. Bacteriol. *188*, 2865–2874.

Perez-Miller, S.J., and Hurley, T.D. (2003). Coenzyme isomerization is integral to catalysis in aldehyde dehydrogenase. Biochemistry *42*, 7100–7109.

Petit, E., LaTouf, W.G., Coppi, M. V., Warnick, T.A., Currie, D., Romashko, I., Deshpande, S., Haas, K., Alvelo-Maurosa, J.G., Wardman, C., et al. (2013). Involvement of a Bacterial Microcompartment in the Metabolism of Fucose and Rhamnose by *Clostridium phytofermentans*. PLoS One *8*, e54337

Petit, E., Coppi, M. V., Hayes, J.C., Tolonen, A.C., Warnick, T., Latouf, W.G., Amisano, D., Biddle, A., Mukherjee, S., Ivanova, N., et al. (2015). Genome and Transcriptome of Clostridium phytofermentans, Catalyst for the Direct Conversion of Plant Feedstocks to Fuels. PLoS One *10*, e0118285.

Pitts, A.C., Tuck, L.R., Faulds-Pain, A., Lewis, R.J., and Marles-Wright, J. (2012). Structural insight into the *Clostridium difficile* ethanolamine utilisation microcompartment. PLoS One *7*, e48360.

Price, G.D., Badger, M.R., Woodger, F.J., and Long, B.M. (2008). Advances in understanding the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. J. Exp. Bot. *59*, 1441–1461.

Purcell, O., Grierson, C.S., Bernardo, M. di, and Savery, N.J. (2012). Temperature dependence of ssrA-tag mediated protein degradation. J. Biol. Eng. *6*, 10.

Quin, M.B., Perdue, S.A., Hsu, S.Y., and Schmidt-Dannert, C. (2016). Encapsulation of multiple cargo proteins within recombinant Eut nanocompartments. Appl. Microbiol. Biotechnol. *100*.

Quin, M.B., Wallin, K.K., Zhang, G., Schmidt-Dannert, C., Donghia, N., Burrill, D., Ferrante, T., McSorley, F.R., Furuta, Y., Vernet, A., et al. (2017). Spatial organization of multi-enzyme biocatalytic cascades. Org. Biomol. Chem. *22*, 185–195.

Rae, B.D., Long, B.M., Badger, M.R., and Price, G.D. (2013). Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO_2 fixation in cyanobacteria and some proteobacteria. Microbiol. Mol. Biol. Rev. *77*, 357–379.

Rondon, M.R., Kazmierczak, R., and Escalante-Semerena, J.C. (1995). Glutathione is required for maximal transcription of the cobalamin biosynthetic and 1,2-propanediol utilization (cob/pdu) regulon and for the catabolism of ethanolamine, 1,2-propanediol, and propionate in *Salmonella typhimurium* LT2. J. Bacteriol. *177*, 5434–5439.

Roof, D.M., and Roth, J.R. (1992). Autogenous regulation of ethanolamine utilization by a transcriptional activator of the eut operon in *Salmonella typhimurium*. J. Bacteriol. *174*, 6634–6643.

Rose, A.S., Bradley, A.R., Valasatava, Y., Duarte, J.M., Prlić, A., and Rose, P.W. (2018). NGL viewer: web-based molecular graphics for large complexes. Bioinformatics.

Sabet-Azad, R., Linares-Pastén, J.A., Torkelson, L., Sardari, R.R.R., and Hatti-Kaul, R. (2013). Coenzyme A-acylating propionaldehyde dehydrogenase (PduP) from *Lactobacillus reuteri*: Kinetic characterization and molecular modeling. Enzyme Microb. Technol. *53*, 235–242.

Salis, H.M., Mirsky, E.A., and Voigt, C.A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. Nat. Biotechnol. *27*, 946–950.

Sampson, E.M., and Bobik, T.A. (2008). Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. J. Bacteriol. *190*, 2966–2971.

Sarrion-Perdigones, A., Falconi, E.E., Zandalinas, S.I., Juárez, P., Fernández-del-Carmen, A., Granell, A., and Orzaez, D. (2011). GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. PLoS One *6*, e21622.

Seedorf, H., Fricke, W.F., Veith, B., Brüggemann, H., Liesegang, H., Strittmatter, A., Miethke, M., Buckel, W., Hinderberger, J., Li, F., et al. (2008). The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. Proc. Natl. Acad. Sci. U. S. A. *105*, 2128–2133.

Shively, J.M., Ball, F., Brown, D.H., and Saunders, R.E. (1973). Functional Organelles in Prokaryotes: Polyhedral Inclusions (Carboxysomes) of Thiobacillus neapolitanus. Science (80-.). *182*, 584–586.

Shone, C.C., and Fromm, H.J. (1981). Steady-state and pre-steady-state kinetics of coenzyme A-linked aldehyde dehydrogenase from *Escherichia coli*. Biochemistry *20*, 7494–7501.

Sievers, F., and Higgins, D.G. (2014). Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol. Biol. *1079*, 105–116.

Slininger Lee, M.F., Jakobson, C.M., and Tullman-Ercek, D. (2017). Evidence for Improved Encapsulated Pathway Behavior in a Bacterial Microcompartment through Shell Protein Engineering. ACS Synth. Biol. *6*, 1880–1891.

Sriramulu, D.D., Liang, M., Hernandez-Romero, D., Raux-Deery, E., Lünsdorf, H., Parsons, J.B., Warren, M.J., and Prentice, M.B. (2008). *Lactobacillus reuteri* DSM 20016 produces cobalamin-dependent diol dehydratase in metabolosomes and metabolizes 1,2-propanediol by disproportionation. J. Bacteriol. *190*, 4559–4567.

Stojiljkovic, I., Bäumler, A.J., and Heffron, F. (1995). Ethanolamine utilization in *Salmonella typhimurium*: nucleotide sequence, protein expression, and mutational analysis of the cchA cchB eutE eutJ eutG eutH gene cluster. J. Bacteriol. *177*, 1357–1366.

Takenoya, M., Nikolakakis, K., and Sagermann, M. (2010). Crystallographic insights into the pore structures and mechanisms of the EutL and EutM shell proteins of the ethanolamine-utilizing microcompartment of *Escherichia coli*. J. Bacteriol. *192*, 6056–6063.

Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P. V., Moriarty, N.W., Zwart, P.H., Hung, L.-W., Read, R.J., Adams, P.D., and IUCr (2008). Iterative model building, structure refinement and density modification with the *PHENIX AutoBuild* wizard. Acta Crystallogr. Sect. D Biol. Crystallogr. *64*, 61–69.

Tsai, S.J., and Yeates, T.O. (2011). Bacterial microcompartments insights into the structure, mechanism, and engineering applications. Prog. Mol. Biol. Transl. Sci. *103*, 1–20.

Tuck, L.R., Altenbach, K., Ang, T.F., Crawshaw, A.D., Campopiano, D.J., Clarke, D.J., and Marles-Wright, J. (2016). Insight into Coenzyme A cofactor binding and the mechanism of acyl-transfer in an acylating aldehyde dehydrogenase from *Clostridium phytofermentans*. Sci. Rep. *6*, 22108.

Wang, B., Kitney, R.I., Joly, N., and Buck, M. (2011). Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. Nat. Commun. *2*, 508.

Wang, M.-F., Han, C.-L., and Yin, S.-J. (2009). Substrate specificity of human and yeast aldehyde dehydrogenases. Chem. Biol. Interact. *178*, 36–39.

Weber, E., Gruetzner, R., Werner, S., Engler, C., and Marillonnet, S. (2011). Assembly of designer TAL effectors by Golden Gate cloning. PLoS One *6*, e19722.

Werner, S., Engler, C., Weber, E., Gruetzner, R., and Marillonnet, S. (2012). Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. Bioeng. Bugs *3*, 38–43.

Yeates, T.O., Jorda, J., and Bobik, T.A. (2013). The Shells of BMC-Type Microcompartment Organelles in Bacteria. J. Mol. Microbiol. Biotechnol. *23*, 290–299.

Zarzycki, J., Erbilgin, O., and Kerfeld, C.A. (2015). Bioinformatic characterization of glycyl radical enzyme-associated bacterial microcompartments. Appl. Environ. Microbiol.

Zarzycki, J., Sutter, M., Cortina, N.S., Erb, T.J., and Kerfeld, C.A. (2017). In Vitro Characterization and Concerted Function of Three Core Enzymes of a Glycyl Radical Enzyme - Associated Bacterial Microcompartment. Sci. Rep. *7*, 42757.

Zuroff, T.R., and Curtis, W.R. (2012). Developing symbiotic consortia for lignocellulosic biofuel production. Appl. Microbiol. Biotechnol. *93*, 1423–1435.